Marijn Janssen · Matti Mäntymäki
Jan Hidders · Bram Klievink
Winfried Lamersdorf · Bastiaan van Loenen
Anneke Zuiderwijk (Eds.)

LNCS 9373

Open and Big Data Management and Innovation

14th IFIP WG 6.11 Conference on
e-Business, e-Services, and e-Society, I3E 2015
Delft, The Netherlands, October 13–15, 2015, Proceedings

Springer

Lecture Notes in Computer Science 9373

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

More information about this series at http://www.springer.com/series/7407

Marijn Janssen · Matti Mäntymäki
Jan Hidders · Bram Klievink
Winfried Lamersdorf · Bastiaan van Loenen
Anneke Zuiderwijk (Eds.)

Open and Big Data Management and Innovation

14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015
Delft, The Netherlands, October 13–15, 2015
Proceedings

Springer

Editors
Marijn Janssen
Faculty of Technology, Policy and Management
Delft University of Technology
Delft
The Netherlands

Matti Mäntymäki
Turku School of Economics
University of Turku
Turku
Finland

Jan Hidders
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Delft
The Netherlands

Bram Klievink
Faculty of Technology, Policy and Management
Delft University of Technology
Delft
The Netherlands

Winfried Lamersdorf
Department of Informatics
Hamburg University
Hamburg
Germany

Bastiaan van Loenen
The Faculty of Architecture and the Built Environment
Delft University of Technology
Delft
The Netherlands

Anneke Zuiderwijk
Faculty of Technology, Policy and Management
Delft University of Technology
Delft
The Netherlands

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-25012-0 ISBN 978-3-319-25013-7 (eBook)
DOI 10.1007/978-3-319-25013-7

Library of Congress Control Number: 2015950044

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Since its inception in 1998, the I3E Conference has brought together researchers and practitioners from all over the world. The I3E conference main area is in the field of e-business, e-services, and e-society, which are abbreviated as the three Es. The I3E conference series is truly multidisciplinary covering areas ranging from computer science to information systems and data science. The 2015 I3E conference was the 14th consecutive I3E conference. This year's conference received submissions from Asia, Europe, and South America.

The mission of IFIP Working Group 6.11 is to organize and promote exchange of information and co-operation related to all aspects of electronic business, electronic services, and electronic society (the three Es). Working Group 6.11 has members with diverse backgrounds including information systems, computer science, and business studies. The I3E conference series the is flagship event of the Working Group.

The theme of the 2014 I3E conference was "Open and Big Data Management and Innovation" Information economy and openness have been the dominating terms over the last couple of years. Data have become widely available and many innovations are based on the utilization of data. The Internet of Things (IoT) enables the availability of large volumes of data, while people create data using social media. Today we have a worldwide exchange of data and this influences our society, resulting in new business opportunities and new services. This has already resulted in the rise of "data science," which aims at better understanding how to use all these data in our information society and how it can co-evolve with its information and communication technology (ICT). This needs to make the theory of complex systems applicable to the information society.

Furthermore, there is a lack of tools and instruments to deal with the vast amount of data. Big data will have a major influence on the operating of businesses and the functioning of society. This conference contributes to advancements in areas such as data analytics, mining, visualization, sensor networks, information retrieval, and information extraction; research relating to data extraction and analytics, statistical inference, data quality, and issues such as bias, missing data, endogeneity, user interface, and visualization are encouraged. Papers are in the field of technology, business, society, or combinations thereof.

The host of the IFIP I3E conference was the Faculty of Technology, Policy and Management (TPM) at Delft University of Technology, The Netherlands. TU Delft is the largest and most comprehensive university of engineering sciences in The Netherlands. The I3E 2015 conference was organized in cooperation between three faculties of Delft University of Technology. The faculties of Technology, Policy and Management (TPM), The Faculty of Architecture and the Built Environment, and the Faculty of Electrical engineering, Mathematics and Computer Science (EEMS). Big and open data (BOLD) is an important topic at Delft University of Technology and there are two main initiatives within Delft University of Technology in this area. Delft

Data Sciences (DSS) is a framework initiative for research activities in data science (http://www.delftdatascience.tudelft.nl/). DDS is characterized by the focus on the engineering aspects of data science, in line with TU Delft's key strengths. Furthermore there is the Knowledge Center Geoinformation Governance, which conducts research on institutional, legal, and organizational aspects of sharing and (re)using geographical information (http://www.bk.tudelft.nl/en/about-faculty/departments/otb-research-for-the-built-environment/knowledge-centre-geoinformation-governance/).

Making a successful conference requires resources and commitment. We would like to thank the authors for their submissions. We also wish to thank the reviewers for ensuring the academic standard of the conference. We wish to extend our thanks to Jo-Ann Karna, Laura Bruns, Diones Supriana, and Martijn Milikan. Finally, we would like to thank everyone involved in organizing the conference.

Enjoy reading!

August 2015

Marijn Janssen
Matti Mäntymäki
Jan Hidders
Bram Klievink
Winfried Lamersdorf
Bastiaan van Loenen
Anneke Zuiderwijk

Organization

Conference Chairs

Marijn Janssen	Delft University of Technology, The Netherlands
Jan Hidders	Delft University of Technology, The Netherlands
Bastiaan van Loenen	Delft University of Technology, The Netherlands
Matti Mäntymäki	University of Turku, Finland

Steering Committee

Winfried Lamersdorf	University of Hamburg, Germany
Wojciech Cellary	Poznan University of Economics, Poland
Matti Mäntymäki	University of Turku, Finland

Witness Workshop Chairs

Bram Klievink	Delft University of Technology, The Netherlands
Stefan Henningsson	Copenhagen Business School, Denmark
Yao-Hua Tan	Delft University of Technology, The Netherlands

Local Support

Ricardo Matheus	Delft University of Technology, The Netherlands
Dhata Praditya	Delft University of Technology, The Netherlands
Diones Supriana	Delft University of Technology, The Netherlands
Sélinde van Engelenburg	Delft University of Technology, The Netherlands
Agung Wahyudi	Delft University of Technology, The Netherlands

Organization Chair

Anneke Zuiderwijk	Delft University of Technology, The Netherlands

I3E 2014 Keynote Speakers

Sunil Choenni	Ministry of Security and Justice, The Netherlands
Yogesh Dwivedi	Swansea University, UK
Liesbet van Zoonen	Erasmus University Rotterdam, The Netherlands

I3E 2014 Program Committee

Esma Aimeur	University of Montreal, Canada
Americo Amorim	Daccord SA
Hernán Astudillo	Universidad Tecnica Federico Santa María, Chile
Evandro Baccarin	University of Londrina, Brazil
Aaron W. Baur	ESCP Europe Wirtschaftshochschule Berlin, Germany
Khalid Benali	LORIA - Université de Lorraine, France
Salima Benbernou	Université Paris Descartes, France
Djamal Benslimane	Lyon University, France
Markus Bick	ESCP Europe Wirtschaftshochschule Berlin, Germany
Peter de Bruyn	University of Antwerp, Belgium
Wojciech Cellary	Poznan University of Economics, Poland
Antonio Cerone	IMT Institute for Advanced Studies Lucca, Italy
Narciso Cerpa	University of Talca, Chile
François Charoy	Université de Lorraine - LORIA – Inria, France
Caspar Chorus	Delft University of Technology, The Netherlands
Joep Crompvoets	University of Leuven, Belgium
Tomi Dahlberg	University of Turku/Turku School of Economics, Finland
Bruno Defude	Télécom sudparis, France
Georgios Doukidis	Greece
Christos Douligeris	University of Pireaus, Greece
Enrico Ferro	Istituto Superiore Mario Boella, Italy
Mila Gasco	ESADE, Spain
Yiwei Gong	Wuhan University, China
Adam Grzech	Wroclaw University of Technology, Poland
Mohand-Said Hacid	Université Claude Bernard Lyon 1 - UCBL
Teresa M. Harrison	State University of New York at Albany, USA
Jan Hidders	Delft University of Technology, The Netherlands
Geert-Jan Houben	Delft University of Technology, The Netherlands
Alexandru Iosup	Delft University of Technology, The Netherlands
Marijn Janssen	Delft University of Technology, The Netherlands
Jonna Järveläinen	Turku School of Economics, Finland
Atsushi Kanai	Hosei University, Japan
Ranjan Kini	Indiana University Northwest, USA
Bram Klievink	Delft University of Technology, The Netherlands
Irene Krebs	University of Technology Cottbus, Germany
Winfried Lamersdorf	University of Hamburg Germany
Hongxiu Li	Turku School of Economics, Finland
Yong Liu	Aalto University, Finland
Euripidis Loukis	University of the Aegean, Greece
Luis Felipe Luna-Reyes	University at Albany, USA
José Machado	Centro ALGORITMI, University of Minho, Portugal
Matti Mäntymäki	Turku School of Economics, Finland
Tiziana Margaria	University of Limerick and Lero, Republic of Ireland

Jani Merikivi	Aalto University, Finland
José María Moreno-Jiménez	Universidad de Zaragoza, Spain
Haralambos Mouratidis	University of Brighton, UK
Jose Neves	Universidade do Minho, Spain
Andreas Oberweis	University of Karlsruhe, Germany
Adegboyega Ojo	National University of Ireland Galway, Republic of Ireland
Hervé Panetto	CRAN, University of Lorraine, CNRS, France
Peter Parycek	Danube-University Krems, Austria
Vassilios Peristeras	European Commission, DG DIGIT, ISA Unit, Belgium
Olivier Perrin	Nancy University, France
Nineta Polemi	University of Piraeus, Greece
Reinhard Riedl	Bern University of Applied Sciences, Switzerland
Jarogniew Rykowski	the Poznan University of Economics, Poland
Jari Salo	University of Oulu, Finland
Florence Sedes	I.R.I.T., P. Sabatier University, France
Santosh Shrivastava	Newcastle University, UK
Zhaohao Sun	PNG University of Technology, Lae, PNG, China
Reima Suomi	University of Turku, Finland
Yao-Hua Tan	Delft University of Technology, The Netherlands
Farouk Toumani	Limos, Blaise Pascal University, Clermont-Ferrand, France
Basitiaan van Loenen	Delft University of Technology, The Netherlands
Herve Verjus	Universite Savoie Mont-Blanc, France
Hans Weigand	Tilburg University, The Netherlands
Robert Woitsch	BOC Asset Management
Hiroshi Yoshiura	University of Electro-Communications, Japan
Yanchun Zhang	Victoria University, Australia
Xianfeng Zhang	China
Jing Zhao	China University of Geosciences, China
Anneke Zuiderwijk	Delft University of Technology, The Netherlands

Contents

Big and Open Data

E-Business, E-Services and E-Society

Keynote Panel

Driving Innovation Using Big Open Linked Data (BOLD) Panel

Yogesh K. Dwivedi[1(✉)], Vishanth Weerakkody[2], Marijn Janssen[3], Jeremy Millard[2,4], Jan Hidders[5], Dhoya Snijders[6], Nripendra P. Rana[1], and Emma L. Slade[1]

[1] School of Management, Swansea University, Swansea, UK
{y.k.dwivedi,N.P.Rana,e.l.slade}@swansea.ac.uk
[2] Business School, Brunel University London, Uxbridge, UK
Vishanth.Weerakkody@brunel.ac.uk, jeremy.millard@3mg.org, Jeremy.Millard@brunel.ac.uk
[3] Section of Information & Communication Technology, Faculty of Technology, Policy, and Management, Delft University of Technology, Delft, The Netherlands
M.F.W.H.A.Janssen@tudelft.nl
[4] Danish Technological Institute, Aarhus, Denmark
[5] Web Information Systems Group, The Software and Computer Technology Department, Delft University of Technology, Delft, The Netherlands
A.J.H.Hidders@tudelft.nl
[6] Stichting Toekomstbeeld der Techniek (STT), The Dutch Study Center for Technology Trends, Den Haag, The Netherlands
snijders@stt.nl

Abstract. Governments have always retained public service data internally in their own systems with only limited information provided to the public and other stakeholders such as the business, charitable and NGO communities. However, the rapid advancement of ICTs coupled with electronic publishing via the Internet in the last decade in particular has enabled governments to exploit the potential of wider distribution and use of such data previously held in internal systems. The panellists will discuss how Big, Open and Linked Data (BOLD) can be utilized to drive innovation and what obstacles and challenges may be encountered. Empowering citizens, potential mis-use in identity theft, policy manipulation or market distortion, and the need to combine open data with closed sources will be discussed.

Keywords: Big data · Open data · BOLD · Panel

1 Introduction

Traditionally, governments have retained public service data internally in their own systems and only a limited amount of information was provided to the public and other stakeholders such as the business, charitable and NGO communities. However, the rapid advancement of ICTs coupled with electronic publishing via the Internet in the last

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 3–9, 2015.
DOI: 10.1007/978-3-319-25013-7_1

decade in particular has enabled governments to exploit the potential of wider distribution and use of such data previously held in internal systems. Recognising these developments, a European Union directive to encourage greater realisation of the economic value of public data through its reuse entered into force in 2003, thus, paving the way for governments to open up to the public previously closed data. This sought to both encourage provision and regulate licensing and charging for information by Member States. In 2004, the OECD recommended that all publicly funded research data should be made available.

The dual motivations of transparency of government and the economic potential of the reuse of data have since motivated politicians to promote the adoption of "open data" as a practice for their governments. In the US, the Presidential Open Government Directive in December 2009 required the use of open formats by federal US agencies, and in May 2010 the UK Prime Minister set out plans for opening up government data through the data.gov.uk website. Subsequently, the European Commission published a Communication on Open Data in 2011, and in the same year the USA, UK and initially six other countries were signatories to the Open Government Declaration.

Although selected public sector data has always been available to public bodies and other organisations, albeit with costs and restrictions, progressive moves by governments in countries such as the UK and US to improve availability and ease of reuse (through machine-readability and technical standards) has removed many such barriers. One of the unforeseen effects of the open data movement has been to make more data easily accessible to other actors in the policy space, including researchers, think-tanks and, most significantly, other parts of the public sector and governmental systems, including local governments. However, the availability of open data relating to various public services has led to more transparency around these services and allowed the general public to hold government departments accountable. This has encouraged stakeholders (including citizens, businesses, charities and NGOs) to take an active interest in the way services are currently delivered and stimulated thinking around how to improve services. Yet the actual use of open data is cumbersome and each stakeholder has to do it themselves as there are no proven solutions facilitating the process of using open data for services transformation or co-creation. Moreover, although the open data movement has gathered much momentum in Europe, the majority of open data repositories (e.g. data.gov.uk, etc.) are too generic and not of direct interest to individual citizens.

Although big or large volumes of raw open data when published in an electronic format is machine readable and can be shared online and re-used, on its own open data offers limited potential for decision making. However, when open data is linked (Big, Open and Linked Data – BOLD) with extra information that provides more context, this offers greater opportunities for stakeholders to exploit the data for innovative purposes, for example through collaboration and co-creation. BOLD could also increase the reach of statistical and operational information, and deeper analysis of outcomes and impacts. Indeed, BOLD offers an opportunity to discover new ways of assessing policy and service outcomes, healthcare and wellbeing, measuring development and making appropriate interventions through innovative solutions where needed.

Although the use of BOLD to date remains limited and at an early stage there are many examples that show its potential. When, in 2011, the UK's mapping authority

released all its geodata rather than selling it, the organization completely transformed its business model from being simply a purveyor of raw data for others to use to becoming its customers' partner in collecting, analysing, tailoring, visualizing and applying linked data in meeting their needs. This new value adding and problem solving role boosted and diversified the authority's income. In another example, Washington DC was one of the first cities to make BOLD available, running hackathons and competitions as part of its 'Apps for Democracy' initiative, which eventually saved the City authorities $2.3 million at a cost of $50,000.

In responding to the theme of The 14th IFIP Conference on e-Business, e-Services and e-Society (I3E 2015), the panellists will discuss how Big, Open and Linked Data (BOLD) can be utilised to drive innovation and what obstacles and challenges may be encountered regarding this. The discussion will be based on empirical evidence, using insights from cases; there will be a focus on analytical models, building on conceptual and theoretical arguments, and our aim is to offer directions for future research.

2 Programme

Professor Yogesh Dwivedi will open the panel by presenting an overview of the topic and subsequently will moderate the panel discussion.

Professor Vishanth Weerakkody will evaluate the role of stakeholders in the BOLD arena and discuss some of the challenges facing public agencies and potential users in exploiting public open data. The discussion will particularly centre on the Hype Vs Reality in the BOLD debate and use several European Commission funded R&D projects to discuss the present technical and user problems faced by local governments who are engaged in promoting BOLD for improved decision making. The discussion will also examine how universities, professional institutes and organisations are preparing their graduates and professionals to deal with the BOLD challenge. In the public service space, who are the main users of big and open data? Are citizens interested or prepared to exploit open data? Are conventional statistical tools equipped to deal with BOLD? Which tools are emerging as the front runners for exploiting BOLD? These are some of the other questions that will be debated.

Professor Marijn Janssen will argue that BOLD can be used to empower citizens and to create an ecosystem in which governments, businesses and citizens can strengthen each other. Yet driving innovations from data is a complex process in which both the available data and the users' demands need to be taken into account. The name "data-driven" already implies that data instead of customers' needs are leading. Furthermore all stakeholders have different interests and concerns. Policy-makers want to use the results for their policy-making, businesses want to create new services to enrich their existing products and services, whereas citizens want to know what is going on and have free access to services without having to up their privacy. For this purpose platforms creating a community, providing an overview of the data and collaborative development environment to develop new ideas, are needed. However, the opening and use of data

might make apparent the low data quality of sources and the focus might be on developing tools and platforms without having the user in mind.

Jeremy Millard's discussion will focus on two main but related issues. First, one of the wicked challenges of BOLD is its potential mis-use, either through negligence, corruption or criminality resulting, for example, in identity theft, policy manipulation or market distortion. Although this is a battle that can never be won, what are the technical, legal, regulatory and ethical challenges we face and how might they be addressed, also taking account of both the Snowden and Assange cases? Second, BOLD becomes even more interesting and disruptive when public data is mixed with data from other sources, including, for example, the private data of businesses or citizens. The smart disclosure approach starts from the premise that people, when given access to data and useful decision tools built for example by governments, can use both their own personal data disclosed by them together with other appropriate data. This can be used to make decisions about their own lives, such as healthcare choices, as well as to self-regulate and be able to hold governments and other actors to account, as well as to cooperate and engage with them.

Jan Hidders will discuss the technological data management challenges of BOLD. In order for BOLD to become effective and usable there must be tools that allow ordinary users, people interested in using the data, to select, transform, combine and visualise the data in an ad-hoc fashion. In some sense this has always been the goal of data management tools, but with the advent of BOLD new challenges have been added such as the scale of the data (in number of data sources, sizes of schemas as well as sizes of actual data content), the possible lack or presence of structure and semantics, and the large heterogeneity of the data sources. This puts all kinds of new demands on the existing data management tools that are currently not being met. These demands are not only in terms of scalability and ability to deal with the semantics of the data, but also in terms of interfaces and languages that allow users to deal with such data in a more intuitive way than now is possible. A key question is what underlying data model these tools should be based upon? Should it be XML, JSON or perhaps RDF which comes with semantic abilities? Or should we perhaps go back to classical models such as the relational model? These and other issues will be discussed in this section.

Dhoya Snijders will focus on the ways in which BOLD is altering how citizens and government relate to each other. Already, BOLD pleads for new types of civil servant with different skills, different educations and, some argue, a different epistemological standpoint. The speed with which data can be created and with which correlations within the data can be made is impacting government's deep-rooted quest for causality. Similarly, BOLD is creating new types of citizen, who are data-focused and data-driven. Citizens are increasingly connecting, measuring, analysing and testing public data themselves. The relations between government and citizens are hereby mediated by BOLD. And, as BOLD is itself becoming intelligent, data is not only opened up, linked, and analysed by human actors - machine-learning is picking up speed and quality. Both citizens and governments will increasingly have to deal with non-human actors in the form of intelligent data-driven systems. To do so we need to develop what sociologists

have dubbed double contingency in which we as humans need to understand how intelligent machines will understand us and vice versa.

Nripendra Rana and Emma Slade will assist in organisation and coordination of the panel and note taking of the panel content.

3 Panellist Bios

Yogesh K Dwivedi is a Professor of Digital and Social Media and Head of Management and Systems Section (MaSS) in the School of Management at Swansea University, Wales, UK. His research interests are in the area of Information Systems (IS) including the adoption and diffusion of emerging ICTs (e.g. broadband, RFID, e-commerce, e-government, m-commerce, m-payments, m-government) and digital and social media marketing. His work on these topics has been published (more than 100 articles) in a range of leading academic journals including: CACM, EJIS, IJPR, GIQ, JORS, ISF, ISJ and IJICBM. He has co-edited more than ten books on technology adoption, e-government and IS theory and had them published by international publishers such as Springer, Routledge, and Emerald. He acted as co-editor of fourteen special issues; has organised tracks, mini-tracks and panels in leading conferences; and served as programme co-chair of IFIP WG 8.6 Conference at the prestigious IIM Bangalore, India in 2013. He is Associate Editor of GIQ, EJM and EJIS, Assistant Editor of JEIM and TGPPP, Senior Editor of Journal of Electronic Commerce Research and member of the editorial board/review board of several journals. He is a life member of the IFIP WG8.6 and 8.5.

Vishanth Weerakkody is Professor of Digital Governance in the Business School of Brunel University in London, United Kingdom. His research experience is focused in the area of public sector service transformation through technology. Professor Weerakkody has published more than 100 peer reviewed articles and has guest edited several special editions of journals and books on this theme. He is co-founder of the e-Government track at the Americas Conference on Information Systems (AMCIS) and Transforming Government Workshop at Brunel University, London. Vishanth has many years of experience in R & D projects in the area of digital governance and is currently an investigator in several European Commission and Internationally funded projects on the use of ICTs in the public sector (e.g. LiveCity, OASIS, DAREED, PolicyCompass, UBiPol, CEES, EGovPoliNet, I-MEET, SI-DRIVE). He is the Editor-in-Chief of the International Journal of Electronic Government Research and one of the two international board members of the Digital Government Society of North America. He is a fellow of the UK Higher Education Academy and combines more than 25 years of experience in industry, teaching and research leadership.

Prof. Dr. Marijn Janssen is full Professor in ICT & Governance and head of the Information and Communication Technology section of the Technology, Policy and Management Faculty of Delft University of Technology. His research interests are in the field of orchestration, (shared) services, intermediaries, open and big data and infrastructures within constellations of public and private organizations. He was involved in EU funded

projects in the past (a.o. EGovRTD2020, eGovPoliNet and Engage), is Co-Editor-in-Chief of Government Information Quarterly, Associate Editor of the International Journal of Electronic Business Research (IJEBR), Electronic Journal of eGovernment (EJEG), International Journal of E-Government Research (IJEGR), is conference chair of IFIP EGOV2015 and IFIP I3E2015 conference (about big and open data innovation) and is chairing mini-tracks at the DG.o, ICEGOV, HICCS and AMCIS conferences. He was ranked as one of the leading e-government researchers in a survey in 2009 and 2014 and has published over 320 refereed publications. More information: www.tbm.tudelft.nl/marijnj.

Jeremy Millard is a Senior Research Fellow at Brunel University (London) and Senior Policy Advisor, Danish Technological Institute (Denmark). He has forty years' global experience working with governments, development agencies, and private and civil sectors in all parts of the world. In the last twenty years he has focused on how new technical and organisational innovations transform government and the public sector. Work with the European Commission includes research and studies on eGovernment, administrative burden reduction, and on developing business models for ICT and ageing. He also recently led an impact assessment of the European eGovernment Action Plan, led the large-scale Europe-wide survey and analysis of eGovernment eParticipation, and developed the 2020 Vision Study on Future Directions of Public Service Delivery. Jeremy has also worked since 2008 as an expert for the UN on their successive global eGovernment development surveys, and provided inputs to both the World Bank and the OECD on eGovernment developments, for example through a survey on back-office developments in support of user-centred eGovernment strategies, as well as ICT-enabled public sector innovation. He works extensively outside Europe in these and related areas, including in the Gulf, the Western Balkans, Georgia, Japan, and India.

Dr Jan Hidders is Assistant Professor in the Web Information Systems group of the Software and Computer Technology department of Delft University of Technology. His research interests are in data integration, data indexing, data linking and large scale data processing, particularly in the domain of graph processing. He was involved in several EU projects such as GRAPPLE and ImREAL where he worked on data integration in the e-learning domain. He has published over 60 refereed publications in conferences such as ICDE, CCGRID, ICDT, CIKM, IUI, FoIKS, ICWE and ISWC, and in journals such as Information Systems, Fundamenta Informaticae, Theory of Computing Systems, BPM Journal and Journal of Computer and System Sciences. He has edited special issues in Fundamenta Informaticae and the Journal of Computer and System Sciences. He has co-organised several workshops such as SWEET (Scalable Workflow Enactment Engines and Technology) at the SIGMOD conference and BeyondMR (Beyond MapReduce) at the EDBT conference. He is currently one of the co-organizers of the 2015 EDBT Summer School on Graph Data Management.

Dr Dhoya Snijders works for STT, the Dutch Study Center for Technology Trends and is currently carrying out a technology foresight study on big data. He holds an MA in Philosophy from the University of Amsterdam and an MA and PhD in Organizational Sciences and Public Administration from the VU University. His research mainly

focuses on classifications and their consequences. He worked as a consultant for some years in the field of e-Government within the Dutch Ministries of Healthcare and Justice. There he carried out national and international projects on ICT governance, eHealth, Open Data, and Big Data. During this time he published on the implementation of ICT in a context of multilevel governance.

Dr Nripendra P Rana is a Lecturer at the School of Management at Swansea University in the UK. He holds a BSc in Mathematics (Hons), an MCA, an MTech, and an MPhil degree from Indian universities. He also holds an MBA with distinction and a PhD from Swansea University. His current research interest is in the area of technology and e-Government adoption and diffusion. He has published his work in some refereed journals including ISF, ISM, ESJ, IJBIS, IJICBM, IJEGR, and TGPPPP. He has varied work experience of teaching in the area of computer science at undergraduate and postgraduate levels. He also possesses a good experience of software development and leading successful software projects.

Emma Slade is a Research Officer in the School of Management at Swansea University. She holds a BSc (Hons), MSc with distinction, and PhD in Business Management from Swansea University. Emma's research interests include consumer and merchant mobile payment adoption, e-government adoption, and consumer forgiveness. Her research has been published in Psychology & Marketing, Journal of Strategic Marketing, and Journal of Computer Information Systems. She also presented at the 2014 Academy of Marketing Science Conference and the 2013 UK Academy for Information Systems Conference. Emma has been invited to review papers for Information Systems Frontiers, Internet Research, and a number of conferences.

Adoption

Adoption of Mobile Banking in Jordan: Exploring Demographic Differences on Customers' Perceptions

Ali Abdallah Alalwan[1], Nripendra P. Rana[2], Yogesh K. Dwivedi[2(✉)], Banita Lal[3], and Michael D. Williams[2]

[1] Amman College of Banking and Financial Sciences, Al-Balqa' Applied University, P.O. Box 1705 Amman Jordan
alwan.a.a.ali@gmail.com
[2] School of Management, Swansea University, Swansea, SA2 8PP, UK
{nrananp,ykdwivedi}@gmail.com, m.d.williams@swansea.ac.uk
[3] Nottingham Business School, Nottingham Trent University, Nottingham, NG1 4BU, UK
banita.lal@ntu.ac.uk

Abstract. This study aims to explore whether Jordanian customers' perceptions on intention and adoption of Mobile banking (MB) services varies according to their demographic characteristics. As per the prior literature, five demographic factors, namely age, gender, income, education and customer's experience with computer and Internet have been considered in the current study. The required data were collected from the field survey questionnaires administered to a convenience sample of Jordanian banking customers. The major statistical results (mean and standard deviation) demonstrate that the customers' perceptions on intention and adoption of MB are likely to vary due to customers' demographic differences. According to the current study's findings, it was also noticed that despite the fact that the most of the sample respondents express a high intention to adopt MB, the adoption rate for the majority of MB services was low.

Keywords: Mobile banking · Customer · Jordan · Demographic differences · Adoption · Behavioral intention

1 Introduction

Mobile banking (MB) is identified as the "use of mobile terminals such as cell phones and personal digital assistants to access banking networks via the wireless application protocol (WAP)" [44, p. 760]. Such innovative banking channel has been progressively implemented over the banking context worldwide to launch customers a wide range of the higher quality banking services (i.e. balance enquiries, fund transfers, payment of bills) without any time or place restrictions [42].

In Jordan, banks seem to be more motivated to adopt MB as an essential banking channel to increase the geographical coverage, customer's satisfaction and loyalty as well as to minimise the operational and labour costs related to traditional branches [4, 26, 32, 33, 42]. Further, the mobile and telecommunication area is growing phenomenally; where there are four mobile services providers along with up to 8.984 million of

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 13–23, 2015.
DOI: 10.1007/978-3-319-25013-7_2

mobile subscriptions [37]. Practically, out of 26 banks working in Jordan, 15 banks have Launched MB services [26].

However, banks worldwide express their concerns regarding the lower adoption rates of MB services by customers. By the same token, Jordanian banking customers seem to be not fully motivated to adopt MB [5]. According to recent study by Alafeef et al. (2011), 6 % of the total Jordanian banking customers have actually used MB services. Alafeef et al. [3] also mentioned that 31 % of banking customers are not aware of the existence of MB services introduced by their banks.

Accordingly, it could be concluded that the main challenge pertaining to successful implementation of MB is conceiving bank clients to move from using human encounter to fully adopt MB [44]. Thus, there is always a necessity to identify the main factors that could hinder or enhance customers' intention and adoption of MB. However, MB related issues have been rarely examined in Jordan. Furthermore, banks need to have more information regarding customers' perception on intention to use and adoption of MB that could differ as per the demographic differences. This, in turn, will help these banks to conduct a useful market segmentation strategy that could accelerate the customers' intention and adoption of such technology [14, 16, 31, 34]. Accordingly, this study aims to conduct an empirical examination to discover if there are considerable variations in the Jordanian customer's perception on the intention and adoption of MB.

2 Literature Review

Theoretically, many studies have endeavored to interpret the most important factors that could hinder or foster customer intention and acceptance of Mobile banking. For instance, Püschel et al. [30] claimed that Brazilian customers' attitudes towards Mobile banking were significantly affected by relative advantage, followed by compatibility; ultimately enriching the customers' intention to adopt Mobile banking. Brown et al. [7] also found that banking customers are more enthused to adopt Mobile banking by relative advantage, trialability, and consumer banking needs. Perceived benefits and governmental regulations have been confirmed as the key positive enablers of customer attitudes towards Mobile banking in Indonesia [29]. In line with this, customers were found to be more motivated to use Mobile banking if they recognised Mobile banking as being useful in their daily life, compatible with their habits and other technologies, and less expensive [18, 42]. By the same token, customer intention to adopt Mobile banking was significantly determined by the role of perceived usefulness, monetary cost, self-efficacy and perceived ease of use [23]. Later, Zhou [43] empirically supported the considerable role of a bank's reputation, information quality, self-efficacy, service quality, and system quality in shaping the customers' initial trust in Mobile banking. More recently, Hanafizadeh et al. [18] all supported the crucial role of perceived usefulness and ease of use in motivating customers to adopt Mobile banking.

With regard to the role of demographic factors, Laukkanen and Pasanen [22] indicated that customers' adoption of Mobile banking was exclusively predicted by the customers' age and gender, but was not predicted by education level, career, family size and income. Riquelme and Rios [35] revealed that the males' intention to adopt Mobile

banking was strongly affected by the role of perceived usefulness, while females paid particular attention to aspects related to ease of use. Chiu et al. [9] indicated that there were no statistical differences in the role of these factors on customers' intention. Laukkanen [21] also discussed that regardless of age differences, customers seem to be more reluctant to adopt Mobile banking due to the negative role of value and usage barriers, while mature customers were observed to be more concerned about barriers pertaining to risk, image, and traditional barriers.

Nevertheless, there is a dearth of literature addressing customer intention and usage of Mobile banking by Jordanian banking customers [5, 19]. Both Khraim et al. [19] and Awwad and Ghadi [5] have found that Mobile banking characteristics - trialability, complexity, compatibility, relative advantages, and risk - are the key predictors of Jordanian customer intention and adoption of Mobile banking. Yet, there is still a need to clarify and empirically examine the important role of the customers' characteristics (i.e. age, gender, income, education, and technology) in shaping the Jordanian customers' perception on intention to use MB and adoption of such emerging systems. Thus, in order to fill this gap, *current study intends to empirically test and explore whether Jordanian customers' perceptions on intention and use of MB could vary according to their demographic characteristics: age, gender, income, education, customers' experience with computer and Internet*.

3 Theoretical Basis

The impact of demographic factors has received a great deal of attention from information systems studies, which assert that variations in customers' reactions to and perceptions of technology could be attributed to the variation in the customers' demographic characteristics [11, 28]. Therefore, this study examines how the demographic factors (such as age, gender, income, education, and customers' experience with computer and Internet) could reflect differences in the Jordanian customers' perception on intention to use and actual adoption of Mobile banking services. These five factors have been widely studied and examined over the online banking area [12, 24, 25]. Further justification and discussion regarding each of these factors are provided in the following subsections.

3.1 Age

Theoretically, age has been debated either as an independent variable or as a factor that could lead a variation on the persons' perception toward a certain kind of individual or collective behaviour and actions [17]. Likewise, prior studies over the information system area have paid a particular interest for the important role of age in shaping the individual perception on the aspects related to technology (i.e. usefulness, ease of use, behavioural intention and actual usage of technology [39–41]. According to Venkatesh et al. [38], actual adopters of computer were found largely to be within the age range of 15 to 35 years. In the self-service technology (SST) context, Dabholkar et al. [11] observed that customers are more likely to be varying on their intention towards and

usage of SST according to their age. Therefore, *it could be expected that the Jordanian banking customers' perceptions on intention and adoption of MB would differ according to their age groups.*

3.2 Gender

According to Morgan [27], gender could be debated either as a descriptive variable or as an explanatory variable. Over the information system (IS) area, a number of authors [39–41] have examined the role of gender in moderating or directly predicting the individual' perception, intention and behaviour toward technology. In keeping with Venkatesh et al. [38], the usage of computer is more likely to be in the higher level among males than females. Differently, some IS researchers [12, 38, 40, 41] have indicated that men are more likely to accept a new technology based on the benefits and advantages perceived; while women usually pay more attention to aspects related to complexity, facilitated resources, and assurance. Therefore, *it could be expected that the Jordanian banking customers' perceptions on intention and adoption of MB would differ according to their gender.*

3.3 Education

In the line with Burgess's [8] proposition, individuals who enjoy an adequate education level are more likely to have a positive perception and ability to conduct a set of complicated actions. Therefore, it has been largely claimed that the well-educated people are more likely to have a positive reaction and perception toward new innovations, thereby; they are more likely to adopt a new technology in comparison with those who are at the less educational level [8, 36]. This thought has been also approved by number of IS studies [2, 28, 38]. For example, Al-Somali et al. [2] empirically proved that Saudi banking customers, who have an adequate level of education, are more likely to have positive attitudes toward Internet banking. Therefore, *it could be expected that the Jordanian banking customers' perceptions on intention and adoption of MB would differ according to their education level.*

3.4 Income

Instead of the employee context where the cost could be restricted in terms of time and effort, customers are more likely to be sensitive to the financial issues that could form their perception toward using the technology [24]. Generally speaking, customers with higher income are more likely to be able carry the financial cost associated with using a new technology [36]. Indeed, using MB services could require customers to pay a cost for using such services in addition to the other cost of having a smart phone, Internet access, and using specified applications [35]. From this perspective, banking customers' perception toward such novel technology could be different according to the customers' income level. This proposition was confirmed by Meuter et al. [25] who empirically approved income level as a key determinant of both customer readiness and customer experiment of different kinds of self-service technologies. Al-Ashban and Burney [1] also empirically approved

income level as a considerable positive predictor of the customers' acceptance of Telebanking in Saudi Arabia. Likewise, according to Kolodinsky et al. [20], the adoption rate of Internet banking channels was observed to be in the higher level among customers who have a higher income level instead of lower income customers. Therefore, *it could be expected that the Jordanian banking customers' perceptions on intention and adoption of MB would differ according to their income level.*

3.5 Experience

Given the particular nature of MB as a self-service banking channel requiring customers to produce the financial services without any assistance from banking staff, the adequate levels of experience and skills with technology could be the important prerequisites to successfully apply this technology (Meuter et al., 2005). Hence, customer experience has been identified by a number of studies as a crucial determinant of customer perception and behaviour towards SST [24, 25]. With reference to Meuter et al. [25], prior experience strongly influences the customer's decision to try a self-service technology, either directly or indirectly, through the mediating impact of customer readiness. Further, customer experience was found to be one of the most influential factors predicting customer propensity toward mobile ticketing in a transportation context [24]. Based on empirical results established by Curran et al. [10] and Chiu et al. [9] customer familiarity in dealing with self-service technology has a significant and positive impact on customers' intention and orientation toward this technology. Therefore, *it could be expected that the Jordanian banking customers' perceptions on the intention and adoption of MB would differ according to customer's experience with computer and Internet.*

4 Research Methodology

Either in information system area or the MB context, it has been highly noticed that the field survey is one of the most prevalent and commonly adopted methods for testing an individual's intention and behaviour towards such an emerging system [13, 15]. Further, this study was conducted with the aim to test and explain the Jordanian banking customers' perception on intention and adoption of Mobile banking. Therefore, the field survey was found to be the most suitable and cost-effective research method allowing access to a large number of Jordanian banking customers in different places within a reasonable time [6]. This, in turn, led to observe that the self-administered questionnaire was a suitable data collection method to obtain the required data from Jordanian banking customers.

As mentioned before, the self-administered questionnaire was selected to derive responses from Jordanian banking customers regarding their perception of the aspects related to behavioural intention and use of Mobile banking. The seven-point Likert scale was used to measure the behavioural intention items with anchors ranging from '1 - strongly agree' to '7 - strongly disagree'. A set of six common financial services was adopted to measure the adoption of Mobile banking. These services have been widely adopted by relevant studies that have examined customers' use or adoption of Internet banking, Mobile banking, and Telebanking [23, 30, 44]. The seven-point time scale was

adopted to measure the use behaviour toward these services with anchors including: 'never', 'once a year', 'several times a year', 'once a month', 'several times a month', 'several times a week', 'several times a day' [41]. Furthermore, six close-ended questions were used for demographic variables such as age, gender, income, education level, Internet experience, and computer experience.

5 Results

5.1 Descriptive Analysis of Usage Behavior

We provide a statistical description regarding the usage patterns of the six Mobile banking services. Balance enquiries and downloading bank statements seem to be the most frequently used Mobile banking services that are applied by the respondents. Indeed, of the 343 valid responses, 105 (30.6 %) used Mobile banking several times per month to look at their bank balance or to download the balance on their bank statement. Yet, balance enquiries and downloading of bank statements via mobile banking have never been used by 85 (24.8 %) of the respondents. In summary, the average mean usage of balance enquiries and downloading bank statements performed by using Mobile banking was 3.65 and the standard deviation was 1.83.

Paying bills was the next widely used Mobile banking service as 110 respondents (i.e. 32.1 %) have used Mobile banking to pay bills once a month while, 130 (37.9 %) valid responses mentioned that they have never used Mobile banking to pay bills. In addition to these, the usage mean of paying bills performed by Mobile banking was 2.77. The third Mobile banking service used by respondents was funds transfer. Even though 145 (42.2 %) of the respondents mentioned that they have never transferred funds through Mobile banking, 118 (34.4 %) [60 + 58] of the respondents indicated that they had utilised Mobile banking for funds transfer about once a month or several times a year. The usage mean of this service was 2.41 and its standard deviation was 1.52. A total of 79 [44 + 35] (23 %) respondents used Mobile banking to request a chequebook or bank certificates several times per year or once a year. Yet, many more respondents (i.e. 179, 52.1 %) have never used the Mobile-banking channel to receive the same services. The mean of using Mobile banking to obtain these services was 2.21 and the standard deviation was 1.52.

Of the few who have used mobile banking for payment of instalments of loans and mortgages, 51 (14.8 %) of the respondents have used Mobile banking for these services once per month. However, the vast majority of respondents (i.e. 224, 65.3 %) reported that they have never applied for these services via Mobile banking. The mean of using mobile banking to conduct these services was too low (about 1.88) with the standard deviation of 1.35.

Moreover, 237 respondents (69 %) have never used these services. Only 94 [33 + 33 + 28] (27.4 %) used Mobile banking to perform these services once a year, several times per year, and once a month respectively. Moreover, these services had the lowest usage mean (i.e. 1.68) among the Mobile banking services and its standard deviation was 1.18.

5.2 Respondent's Demographic Characteristics and Customers' Perception Relating to Intention to Use and Adoption of Mobile Banking

Both mean and standard deviation (SD) are tested in the current study to see how the customers' perception regarding the issues related to behavioral intention and adoption of MB could be vary according to respondent's demographic characteristics. In the term of age difference, all age categories express a high intention to use MB. Yet, according to the average mean accounted for age group 25–30 and 31–40, younger generation seems slightly more interested in using MB in future. The actual adoption mean extracted for age categories indicted the lower adoption rates of MB services by Jordanian customers. However, the adoption rates of MB were noticed in their highest level for age group 25–30 followed by age group of 31–40.

As for gender difference, both male and female seem to be more willing to adopt MB due to the highest mean score accounted in this regard, yet; the mean accounted for male was higher than for female. On the other hand, the usage mean accounted for both male and female was low with value of 2.27 for male and another lower value for female (i.e. 2.012).

According to main statistical outcomes regarding the refection of educational level, as expected, the highest rate of behavioural intention was noticed regarding those respondents who have a PhD degree with mean value of 6.33 followed by those who have masters (5.39) and Bachelors (5.37) degrees. By the same token, the largest mean of the adoption behaviour was accounted for respondents who have a PhD with mean value of 2.28. Lower adoption means were also recorded for the all education groups.

In the term of computer experience, the largest mean value of behavioral intention (i.e. 5.62) was accounted for those respondents who have three years of experience or above with computer while the value of 4.25, which is the lowest value was accounted for those who have less than three years of experience with computer. Similarly, the highest mean value of adoption (i.e. 2.26) was noticed in the case of respondents who have experience of three years and above.

As for Internet experience, the largest mean (i.e. 5.47) of behavioural intention was accounted by those respondents who have experience with Internet for three years and above. On the other hand, the lowest mean (i.e. 4.25) of behavioural intention was in the case of respondents with Internet experience less than one year. By the same token, the largest mean (i.e. 2.22) of adoption of MB services was in the case of Internet experience of three years or above whereas the lowest one (i.e. 1.92) in this regard was in the case of experience group of 1–2 years.

6 Discussion

Generally, the main results extracted in the current study were found to be in line with what has been discussed and approved by prior literature regarding the role of demographic factors on the perception toward technology. For instance, the age categories of 25–30 and 31–40 express a higher willingness toward using MB in comparison with older customers (i.e. 50–60 and those above 61). Even though the adoption rates of MB services are too low over all age groups, younger customers were observed as highly

involved in using MB in comparison to older customers. Such issues could possibly return to the fact that older customers are more likely to not have the important skills and experience that enable them to properly use MB and most of them they do not have smart phones needed to use MB services. Younger customers seem to have more capability and confidence to interact with such sophisticated technology like Mobile banking. This could be regarded to the fact that younger customers have more interaction with technology and Mobile innovation and having an adequate level of technological savvy, awareness, skills, and knowledge, [41, 42].

As for gender, the findings indicated that there are slight differences between males and females in their perception toward intention and adoption. Nevertheless, it could be difficult in the current stage to argue that males are more interested and hence can heavily adopt MB services than females. This could be attributed to the fact that the influence of gender differences are more likely to vanish over the highly evolved communities where both males and females have equal opportunities to be educated, to work, and to get interacted with technologies. In parallel with these results, there are several studies [11, 12] that have empirically disapproved a variation between males and females in their intention and reaction toward technology.

In line with what has been expected regarding the role of education, highly educated respondents seem to be more motivated to use MB and more involved as well in adopting such technology. As discussed earlier in this study, people with a good level of education are more likely to have the sufficient knowledge and skills that could help them to cope with new technology rather than less educated people. Theoretically, this proposition has been highly supported by different studies conducted in the same area of interest [2, 28, 38].

Respondents were also observed to be different in their intention and adoption of MB according to the variation in their income level. While the higher income respondents seem to be more interested and active users of MB services, lower income respondents are less interested and motivated in this regard. Over the marketing literature, it has been highly argued that the higher income customers are less sensitive for the cost issues [25, 36]. Those customers are more able to carry the financial cost associated with buying the important facilities and resources required to use Mobile banking [20].

As it is expected, customer perceptions toward both behavioural intention and adoption are more likely to be different according to the customers' experience with technology and Internet. Indeed, the results regarding the willingness to adopt MB and the actual adoption of MB were able to reach the highest level among respondents who have a good level of technology experience for more than three years. As discussed in Sect. 3.5, the MB channel as self-service technology requires customers to independently conduct all the process to produce and transport the MB services. Therefore, adequate level of experience with technology is a very important aspect to let the customers be more confident in their ability to deal with such complicated systems. These results are consistent with other IS studies, which assured the important role of customers' experience in formulating the customers' perception and reaction to cope with different kinds of systems [10].

This study comprises an important contribution by exploring the main demographic features of adopters and potential adopters of Mobile banking in Jordan as more

emerging system calling for further explanation and examination. Primarily, a theoretical contribution was captured in the current study by synthesising the relevant literature of IS area and Mobile banking as well. Further, such theoretical propositions were empirically tested via collecting sufficient amount of data from the Jordanian banking customers. Practically, the results of the current study alert the Jordanian banks about the current state of the Jordanian customers' intention and adoption of MB services. Therefore, by conducting an empirical study to discover the demographic features of adopters and potential adopters of MB, this study was hoping to provide the Jordanian banks with relevant guidelines that would facilitate an effective implementation and acceptance of MB in proportion to the customer's category and their demographic characteristics. The main results of mean and standard deviation of behavioural intention obviously suggest that the majority of the respondents seem to be more motivated to adopt MB. In addition to this, most of them enjoy an adequate level of education and experience with the Internet and the computer, thus; moving them as actual users of MB will not be expensive and difficult. Therefore, allowing customers to try using these applications through experimental accounts rather than using their own accounts could create a positive experience and let customers actually experience how much they will benefit by using these valuable, useful and easier applications [13].

7 Conclusion

The fundamental intention of the current research was to discover how the Jordanian banking customers' perceptions on aspects related to their intention and adoption of MB could be differing because of their demographic features. Five common demographic variables, namely age, gender, income, education, and technology experience were identified and tested in the current study. The empirical findings clearly suggest that Jordanian banking customers are more likely to be different in their perceptions toward intention and adoption of MB according to their demographic differences.

7.1 Limitations and Future Research Directions

One of the main limitations of the current study is that the data was derived using a convenience sample of banking customers from two cities in Jordan: Amman and Al-Balqa, raising a concern regarding the applicability of the current study results for other banking customers in different regions. Accordingly, it would be more useful for future studies to capture the required data from a large sample size covering the most parts in Jordan. This study only focuses on the customers' demographic features while it does not pay attention to the psychological and behavioral factors (such as habit, innovativeness, customer readiness, and self-efficacy). Therefore, examining such factors along with demographic features could provide a rich understanding of the customer's reaction and perception toward Mobile banking. Future studies should look at the impact of cultural aspects on the Jordanian customers' perceptions toward MB especially when such aspects have not been examined in the area of Mobile banking.

References

1. Al-Ashban, A.A., Burney, M.A.: Customer adoption of Telebanking technology: the case of Saudi Arabia. Int. J. Bank Mark. **13**(5), 191–201 (2001)
2. Al-Somali, S., Gholami, R., Clegg, B.: An investigation into the acceptance of online banking in Saudi Arabia. J. Bus. Res. **29**(2), 130–141 (2009)
3. Alafeef, M., Singh, D., Ahmad, K.: The influence of demographic factors and user interface on mobile banking adoption: a review. J. Appl. Sci. **12**(20), 2082–2095 (2012)
4. Alalwan, A.A., Dwivedi, Y.K., Rana, N.P., Lal, B., Williams, M.D.: Consumer adoption of Internet banking in Jordan: examining the role of hedonic motivation, habit, self-efficacy and trust. J. Financ. Serv. Mark. **20**(2), 145–157 (2015)
5. Awwad, M.S., Ghadi, M.Y.: Investigating of factors influencing the intention to adopt mobile banking services in Jordan. Dirasat Adm. Sci. **37**(2), 545–556 (2010)
6. Bhattacherjee, A.: Social Science Research: Principles, Methods, and Practices, 2nd edn. Florida, USA (2012)
7. Brown, I., Cajee, Z., Davies, D., Stroebel, S.: Cell phone banking: predictors of adoption in South Africa—an exploratory study. Int. J. Inf. Manag. **23**(5), 381–394 (2003)
8. Burgess, R.: Key Variables in Social Investigation. Routledge, London (1986)
9. Chiu, Y., Fang, S., Tseng, C.: Early versus potential adopters: exploring the antecedents of use intention in the context of retail service innovations. Int. J. Retail Distrib. Manag. **38**(6), 443–459 (2010)
10. Curran, J.M., Meuter, M.L., Surprenant, C.F.: Intentions to use self-service technologies: a confluence of multiple attitudes. J. Serv. Res. **5**(3), 209–224 (2003)
11. Dabholkar, P.A., Bobbitt, M.L., Lee, E.J.: Understanding consumer motivation and behavior related to self-scanning in retailing: implications for strategy and research on technology based self-service. Int. J. Serv. Ind. Manag. **14**(1), 59–95 (2003)
12. Dean, D.H.: Shopper age and the use of self-service technologies. Managing Serv. Qual. **18**(3), 225–238 (2008)
13. Dwivedi, Y.K., Irani, Z.: Understanding the adopters and non-adopters of broadband. Commun. ACM **52**(1), 122–125 (2009)
14. Dwivedi, Y.K., Lal, B.: Socio-economic determinants of broadband adoption. Ind. Manag. Data Syst. **107**(5), 654–671 (2007)
15. Dwivedi, Y.K., Choudrie, J., Brinkman, W.P.: Development of a survey instrument to examine consumer adoption of broadband. Ind. Manag. Data Syst. **106**(5), 700–718 (2006)
16. Dwivedi, Y.K., Wastell, D., Laumer, S., Henriksen, H.Z., Myers, M.D., Bunker, D., Elbanna, A., Ravishankar, M.N., Srivastava, S.C.: Research on information systems failures and successes: status update and future directions. Inf. Syst. Front. **17**(1), 143–157 (2015)
17. Finch, J.: Age. In: Burgess, R. (ed.) Key Variables in Social Investigation. Routledge, London (1986)
18. Hanafizadeh, P., Behboudi, M., Koshksaray, A.A., Tabar, M.J.S.: Mobile-banking adoption by Iranian bank clients. Telematics Inform. **31**(1), 62–78 (2014)
19. Khraim, H.S., Shoubaki, Y.E., Khraim, A.S.: Factors affecting Jordanian consumers' adoption of mobile banking services. Int. J. Bus. Soc. Sci. **2**(20), 96–105 (2011)
20. Kolodinsky, J.M., Hogarth, J.M., Hilgert, M.A.: The adoption of electronic banking technologies by US consumers. Int. J. Bank Mark. **22**(4), 238–259 (2004)
21. Laukkanen, T.: Internet vs Mobile banking: comparing customer value perceptions. Bus. Process Manag. J. **13**(6), 788–797 (2007)
22. Laukkanen, T., Pasanen, M.: Mobile banking innovators and early adopters: How they differ from other online users? J. Financ. Serv. Mark. **13**(2), 86–94 (2008)

23. Luarn, P., Lin, H.H.: Toward an understanding of the behavioral intention to use mobile banking. Comput. Hum. Behav. **21**(6), 873–891 (2005)
24. Mallat, N., Rossi, M., Tuunainen, V.: An empirical investigation of mobile ticketing service adoption in public transportation. Pers. Ubiquit. Comput. **12**(1), 57–65 (2008)
25. Meuter, M.L., Bitner, M.J., Ostrom, A.L., Brown, S.W.: Choosing among alternative service delivery modes: an investigation of customer trial of self-service technologies. J. Mark. **69**(2), 61–83 (2005)
26. Migdadi, Y.K.A.: The developing economies' banks branches operational strategy in the era of e-banking: the case of Jordan. J. Emerg. Technol. Web Intell. **4**(2), 189–197 (2012)
27. Morgan, D.H.J.: Gender. In: Burgess, R. (ed.) Key Variables in Social Investigation. Routledge, London (1986)
28. Proença, J., Rodrigues, M.A.: A comparison of users and non-users of banking self-service technology in Portugal. Manag. Serv. Qual. **21**(2), 192–210 (2011)
29. Purwanegara, M., Apriningsih, A., Andika, F.: Snapshot on Indonesia regulations in mobile internet banking users' attitudes. Procedia Soc. Behav. Sci. **115**, 147–155 (2014)
30. Püschel, J., Mazzon, J.A., Hernandez, J.M.C.: Mobile banking: Proposition of an integrated adoption intention framework. Int. J. Bank Mark. **28**(5), 389–409 (2010)
31. Rana, N.P., Dwivedi, Y.K., Williams, M.D.: A meta-analysis of existing research on citizen adoption of e-government. Inf. Syst. Front. **17**(3), 547–563 (2015)
32. Rana, N.P., Dwivedi, Y.K., Williams, M.D., Lal, B.: Examining the success of the online public grievance redressal systems: an extension of the IS success model. Inf. Syst. Manag. **32**(1), 39–59 (2015)
33. Rana, N.P., Dwivedi, Y.K., Williams, M.D., Weerakkody, V.: Investigating success of an e-government initiative: validation of an integrated IS success model. Inf. Syst. Front. **17**(1), 127–142 (2015)
34. Rana, N.P., Dwivedi, Y.K.: Citizen's adoption of an e-government system: validating extended social cognitive theory (SCT). Gov. Inf. Quart. **32**(2), 172–181 (2015)
35. Riquelme, H.E., Rios, R.E.: The moderating effect of gender in the adoption of mobile banking. Int. J. Bank Mark. **28**(5), 328–341 (2010)
36. Rogers, E.M.: Diffusion of Innovations. The Free Press, New York (1995)
37. The Jordan Times.: Mobile phone penetration projected to reach 200 % (2013). http://jordantimes.com/mobile-phone-penetration-projected-to-reach-200 on 2 June 2015
38. Venkatesh, A., Shih, E., Stolzoff, N.: A longitudinal analysis of computing in the home. In: Sloane, A., van Rijn, F. (eds.) Home Informatics and Telematics, pp. 205–215. Springer, US (2000)
39. Venkatesh, V., Morris, M.G.: Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behaviour. MIS Q. **24**(1), 115–140 (2000)
40. Venkatesh, V., Morris, M., Davis, G., Davis, F.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
41. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**(1), 157–178 (2012)
42. Wessels, L., Drennan, J.: An investigation of consumer acceptance of M-banking. Int. J. Bank Mark. **28**(7), 547–568 (2010)
43. Zhou, T.: Understanding users' initial trust in Mobile banking: an elaboration likelihood perspective. Comput. Hum. Behav. **28**(4), 1518–1525 (2012)
44. Zhou, T., Lu, Y., Wang, B.: Integrating TTF and UTAUT to explain mobile banking user adoption. Comput. Hum.Behav. **26**(4), 760–767 (2010)

Young-Elderly Travellers as Potential Users and Actual Users of Internet with Mobile Devices During Trips

Niklas Eriksson(✉) and Susanna Fabricius

Department of Business Management and Analytics,
Arcada University of Applied Sciences, Jan-Magnus Janssonin aukio 1,
00560 Helsinki, Finland
{niklas.eriksson,susanna.fabricius}@arcada.fi

Abstract. The population is rapidly ageing in countries such as Finland. However, little research has been conducted to better understand older travellers' use of Internet and mobile devices. This qualitative study aims at exploring young-elderly (aged 60–75) travellers as potential users and actual users of the Internet with mobile device during trips. The results identify a range of possible drivers and barriers for the use of Internet with mobile devices and their impact on the travel experience during trips. The study also suggests that there is a substantial number of young-elderly travellers that are quite advanced in their mobile usage behaviour.

Keywords: Older travellers · Travel experience · Mobile devices · Technology adoption and use · User behaviour · Digital services

1 Introduction

Many travellers are nowadays acting as their own travel agents and they build their own travel packages and trip itineraries [21]. Mobile technology such as tablet devices and smart phones is taking the digital development even further. A large scale survey by Hjalager and Jensen [10] confirms that many travellers want to be online before the trip, during the trip and after the trip. However, individuals adopt technology based innovations (services) very differently [20]. One consumer segment, which is becoming very relevant, is the so called young-elderly (aged 60–75) segment [1, 4]. Population ageing is faced by most developed countries. For example in Finland the proportion of persons aged 65 or over in the population is estimated to rise from the present 18 % to 26 % by 2030 and to 28 % by 2060 [25]. Even globally the population aged 60 or over is the fastest growing [29]. However, only little research has been conducted on the influence of Internet in the older tourist market [17, 30]. On the other hand an ever growing number of older people in countries such as Finland are Internet users. This appears from Statistics Finland's survey on use of information and communications technology. In 2013, the upper age limit of the survey sample was raised from 74 to 89 years [26]. The number of mobile devices has also quickly increased in the oldest age groups. In fact, the

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 24–35, 2015.
DOI: 10.1007/978-3-319-25013-7_3

adoption rates of smart phones for people aged 60–75 have passed 50 % in Finland [4]. Once older adults join the digital world it tends to become an integral part of their lives [23].

Based on the fact that individuals adopt and use online services differently, the lack of research of older tourists' online behaviour and the increased use of Internet and mobile devices in this segment, it seems necessary and timely to investigate young-elderly travellers' online behaviour. This study will focus on exploring young-elderly travellers as potential users and actual users of Internet with mobile devices during trips.

2 Literature

2.1 Online Activities and Older Travellers

Before the trip travellers are generally focused on assessing the destination and planning transportation and accommodation, whereas in the during trip settings the search strategies of tourists are primarily focused on planning the venue on-site, such as activities to undertake [22]. Seniors research the Internet for travel information in different ways, e.g. through Google searches and by visiting travel sites [27]. However, according to the same authors seniors rarely use social networks such as Facebook and Twitter for travel planning, but many of them have entered an 'e-buyer' era, where they actually buy travel products online. Generally speaking social media is growing in importance as an influence on the online travel information search process [34]. For example travel sites such as Tripadvisor provide global platforms for rating travel services and sharing experiences. Tourists often want to recall memories and share them with others with e.g. photos and stories both during the trip and after the trip [28]. Nowadays numerous travellers come pre-loaded with apps and content for mobile use. Mobile services have the potential to support tourists in different stages of the trip and mobile devices are used for many online travel activities, e.g. search, book and reflect [7]. Many travellers have developed new routines in during trip settings due to the use of a smart phone. These new routines include for example finding information about the travel, book tickets and taking photos and sharing with others immediately [33]. In fact using a personal mobile device for online purposes during trips is seen by many travellers as important [10]. Minazzi and Mauri [14] point out that the use of mobile devices and applications affect the travel experience in different stages of the traveller life cycle. On the other hand Pesonen et al. [17] found that Finnish seniors seldom use their mobile devices to connect to the Internet during trips. Others have also emphasized that we should be careful to overestimate the extent of travellers' use of mobile devices [11]. However, neither should we underestimate the number of technology savvy older travellers. Research conducted by Reisenwitz et al. [19] shows that seniors are online more hours and more frequently and feel more comfortable online. Niemelä-Nyrhinen [15] concluded that elderly (baby boomers) in Finland have, generally speaking, a low level of technology anxiety.

2.2 Drivers and Barriers of Technology Use and Impact on Travel Experience

Several models have been developed to better understand individuals' adoption and use of information technology. One of the most widely used models is the technology acceptance model (TAM) by Davis [6] which is based on the theory of reason action (TRA) by Fishbein et al. [8] with routes in psychology theories and its extension theory of planned behavior (TPB) by Ajzen [2]. Other often used models in technology adoption research are the diffusion of innovations theories (DOI) by Rogers [20], the unified theory for the acceptance and use of technology (UTAUT) by Venkatech et al. [31] and the unified theory for consumers' acceptance and use of information technology (UTAUT2) by Venkatech et al. [32]. These models have been applied to explain the acceptance of different types of technologies by older adults [5]. Mallenius et al. [12] found through the lens of UTAUT, that Expected benefit, Perceived security, Usability, Anxiety, Training, Guidance, Price barriers and Social influence are relevant when investigating elderly individuals' mobile device and service use in Finland.

It should be noted that when investigating adoption and use of technology it is very important to distinguish between *potential users* and *actual users*. Gerpott [9] found that innovation-based attributes explained mobile Internet acceptance better for actual users than for potential users. Recent studies have also emphasized the need for research on the impact of technology on the tourist experience among those who actually uses their mobile devices [16]. The same authors concluded that emotional responses, missed opportunities, monetary burden and behavioural consequences may lead to negative travel experience effects. In fact, Wang et al. [33] emphasized that not only drivers of smart phone use in travel should be examined in order to understand its impact on travel experience, but also barriers (resistance) to use. Experiences of technology while travelling may also induce anxieties and tensions, due to technology addiction [18].

3 Methodology

We conducted a semi-structured qualitative study with 14 pensioners within the age group 60–75 to better understand young-elderly travelers as potential users and as actual users of the Internet with mobile devices during trips. The sample was drawn from a Swedish speaking pensioners' association in Helsinki Finland. In order to gain as broad as possible view of the target segment both low-proficiency users of the Internet and high-proficiency users of the Internet were selected in collaboration with the association. Also low frequency travelers and high frequency travelers were selected. The selection of the sample can hence be described as purposive [13]. All informants were strangers to the two interviewers and the interviews lasted for about an hour on average. The interviews were conducted in spring 2015 and according to ethical standards and confidentiality by providing advance information to the informant regarding the purpose of the study, as well as about who will have access to the data and confidentiality guidelines of the project.

An instrument was developed to guide the interviews. The questions analyzed in this study are listed below:

- Describe how you mostly search for information during your trips
- Do you use the Internet with your smart phone and/or tablet device during your trips?
- What do you do with them? Which device do you prefer?
- Is there a difference in how you use the device abroad compared to your use in Finland?
- Why do you use the Internet with your smart phone and/or tablet device during your trips?
- Why do you not use the Internet with your smart phone and/or tablet device during your trips?
- Do you see that that the use of the Internet with a smart phone and/or tablet device during your trips impacts your travel experience?
- How do you share your travel experiences during your trips?

The questions are flexible in character which allowed us to account for individual differences and take the advantage of the iterative nature of interviewing [13]. The interviews were conducted in Swedish and voice recorded. Direct citations have been translated to English by the authors.

The sample may be biased towards people with a higher educational degree and a larger household income than the average Finnish 60–75 aged individuals (cf. [17]). On the other, hand our purpose was to select as diverse informants as possible based on travel frequency and online proficiency and not based on educational and economic situation. From Table 1 we can see that the informants have a wide distribution in perceived online proficiency, annual travel frequency and numbers of trips during the past three years. Furthermore we are not aiming at generalizing results but to explore possible drivers and barriers for the potential use and actual use of the Internet with mobile devices during trips.

4 Results

According to the theory discussed and based on the results we have divided the sample into *potential users* and *actual users* of the Internet with mobile devices in during trip settings. Five informants (F1, F2, M7, F11 and M13) belong to potential users and nine (F3, F4, M5, M6, M8, F9, F10, M12 and F14) belong to actual users. With the theoretical discussion on technology use and impact on travel experience in mind we analysed the results of the two groups.

4.1 Potential Users

The results are sub-divided into themes that represent the drivers and barriers of potential users of the Internet with mobile devices during trips.

Expected Added-Value. F2, M7 and F11 said that they see themselves using a mobile device, either smart phone or tablet device, during the trip in the near future. They all three see potential added value by using a mobile device during a trip, e.g. fast and convenient access to information, locating themselves on a map. In fact two of them (F11, M7) had already purchased a smart phone, but had not taken it into use. F11 said that the smart phone was purchased specifically for the needs of an upcoming trip

Table 1. Interview informants

Nr	Gender	Age	Retired	Education	Annual household income	Family	Travel Freqency	Trips past 3 years, Domestic	Trips past 3 years, Abroad	Perceived online proficiency	Internet importance during trip	Main purpose to travel	Main travel style	Preferred destination	Main travel partner	Barriers to travel
F1	Female	75	2003	Vocational	12.000	Living alone	Once a year	8	3	Poor	Not at all	Holiday, visit friends, children / grandchildren	Self-organized	New and Known	Spouse	None
F2	Female	69	2008	Vocational	25.000	Living alone	2-7 times a year	15	10	Moderate	Slightly	Senior dance trips	Packaged trips	New	Senior dancers	None
F3	Female	71	2010	University	90.000	Married	2-7 times a year	10	12	Excellent	Very	Holiday, visit friends, children / grandchildren	Self-organized	New and Known	Spouse	None
F4	Female	62	-	Vocational		Widow	2-7 times a year	1	2	Good	Medium	Holiday, Senior dance trips	Packaged trips	New and Known	Friends, Children / grandchildren	None
M5	Male	62	-	University	90.000	Married	2-7 times a year	4	11	Moderate	Very	Holiday, Work	Self-organized	New	Friend	Economy
M6	Male	73	2005	Vocational	100.000	Married	2-7 times a year	10	10	Moderate	Very	Holiday, visit friends, children / grandchildren	Self-organized	Known	Spouse	None
M7	Male	74	2002	Vocational	50.000	Married	>= once/ month	4	3	Good	Not at all	Senior trips	Packaged trips	Known	Pensioneers	None
M8	Male	60	2014	University	35.000	Living alone	2-7 times a year	5	10	Moderate	Medium	Holiday and visit friends	Self-organized	New	Alone	None
F9	Female	60	2014	University	110.000	Married	2-7 times a year	4	7	Moderate	Extremely	Holiday and visit friends	Self-organized	New and Known	Friend and spouse	Economy
F10	Female	72	2003	University	80.000	Married	2-7 times a year	6	3	Mediocre	Very	Holiday and visit friends	Self-organized	New	Friend	Family
F11	Female	69	2006	University	55.000	Married	>= once/ month	25	1	Mediocre	Not at all	Summer house	Self-organized	Known	Alone	Family
M12	Male	69	2013	Vocational	70.000	In relationship	2-7 times a year	10	7	Excellent	Extremely	Holiday and visit friends	Self-organized	New	Spouse	None
M13	Male	72	2005	Vocational	60.000	Married	Once a year	2	4	Excellent	Very	Holiday	Packaged trips	New	Friends, Spouse, Children / grandchildren	Health
F14	Female	71	2008	University	45.000	Widow	2-7 times a year	0	15	Good	Medium	Holiday, visit friends, children / grandchildren	Self-organized	New and Known	Alone. Friends, Children / Grandchildren	Economy

"I need to be able to locate myself on a map during my trip to Spain and therefore I have purchased a smart phone, but I have not taken it into use yet." (F11)

F1 on the other hand found it more challenging to perceive the travel experience enhancement of using smart phones, rather she finds them making people anti-social and distracted from the world around them.

"*I cannot really see the purpose of people sitting in trains, busses etc. with their heads down …. Phones should be used only as phones …"* (F1)

M13 said he owns an old smart phone that he does not use, because it is slow and he does not really need it.

Travel Style and Partner. F1 and F2 see that their travel style influences their behaviours and that they form different roles with their travel partners during the trip. In fact, because their travel partner uses a mobile device they do not need to use one themselves.

"I mostly travel together with my sister… My sister is searching for information with a mobile device during trips so I don't really need to." (F2)
"When we were in Stockholm my friend used her tablet device to find information on where to go and what to do… I do the talking with locals and she looks for information." (F1)

M7, F1 and F2 mostly go on packaged trips (see Table 1). M7 sees that because he mostly takes part in organised senior-trips he does not really need a mobile device to arrange things during the trip. F2 on the other hand sees herself using a personal mobile device when she goes on trips without her sister.

"I probably need to start using one, as I cannot go on every trip with my sister." (F2)

Knowledge and Guidance. Four of the informants (F1, F2, M7 and F11) expressed that technology in general can be struggling and frustrating to use. They admit that their personal technology skills are limited but that it can be hard for them to admit to others that they have problems using technology appliances. M7 said that he has a smart phone waiting for him, but he has not taken it into use. The reason is that he needs help in learning how to use it.

"I have been struggling to make the effort to get it up and running. I should ask help from my daughter so that I will be able to use it." (M7)

Anxiety. Anxiety towards technology does not seem to be a hindrance for F2, F11 and M13 to become actual mobile device users during trips, as they are looking for and booking travel on the Internet in the pre-trip stage. F11 is mostly purchasing routine trips on the Internet (train, ferry and bus tickets) and F2 and M13 have purchased flights, accommodation and travel packages online. M7 is using the Internet for finding travel information but he has not tried to purchase travel online. M7 feels more comfortable using traditional travel agents to make trip arrangements and prefers to go on packaged senior-trips, rather than to organize them himself online.

"I call the local travel agent to make trip arrangements… I can trust them as I have used them a lot before." (M7)

F1 concluded that she is trying to stay away from technology appliances and even her children have told her to stay away from technology. She even considered her-self lucky to be out of working life due to so much now being computer-based and she could not cope with that. Her only point of contact with the Internet is if someone wants her to check some information on a web-site.

4.2 Actual Users

The results are sub-divided into themes that represent actual drivers and barriers of using the Internet with mobile devices during trips and their impact on travel experience.

Added-Value. The informants use Internet with mobile devices during trips to find local sights, check reviews, weather and opening hours, find the shortest routes to places etc. Map services are the most widely used mobile services. The informants find the mobile device convenient and time-saving as they have instant access to information in any situation. F14 said that she is very attached to her iPhone and F9 that it feels like she has inside information about local places as she can check other travelers' experiences online with her mobile device. This kind of information was not available to her before or it was not easily accessible. However, it can also have a negative effect on the travel experience as it may passivate travelers from asking locals etc., making travelers actually missing out on experiences.

> "*…on the other hand mobile devices may passivate us to ask locals, as we start to believe that we already know everything about the place.*" (F9)

In fact, M8 sometimes intentionally leaves his mobile device out and asks e.g. the hotel reception for local tips as they may have some really valuable knowledge to offer. According to M8 visiting the online top rated attractions may not give him the authentic travel experience he is looking for. Also digital map services may be found too efficient.

> *"Nowadays we do not get lost and stumble into interesting things like we used to do."* (F3)

The informants also kill or fill time in transportation by using their mobile devices. Only one informant (M12) mentioned that he uploads pictures with his mobile device in social media during trips. All other informants prefer to share their experiences after the trip. However, social media is not extensively used for sharing travel experiences after the trip either, primarily due to unwillingness to share private information.

Mobile devices also give the informants a feeling of safety, that the traveler can be reached and that they can reach travel partners and people at home (e.g. SMS, e-mail, WhatsApp, Facebook). On the other hand the awareness of things at home may also increase tensions during the trip (see technology tensions below).

Travel Style. All nine informants reported that they mainly organize their trips themselves, rather than taking part in packaged trips (see Table 1). They find the mobile device is a key tool for their style of traveling and experiencing things. The following two narratives represent their responses well.

> *"I want to organize things myself and then the mobile device comes in handy." (M8)*

"*I can imagine that if I would take part in an organized trip I would not really need my mobile device as everything then is organized.*" (F9)

Some of them even feel reluctant to take part in organized senior-trips. It may even be hard for them to see or they don't want to see themselves as seniors by definition.

"That sounds a little bit boring… I think that is for people who cannot travel in another way….it is good that they are arranged but that is not for me." (M8)

Usability. All nine informants found mobile devices to be mostly easy to use and bring along while travelling. Most of them bring both a tablet device and a smartphone with them on the trip. However, there are different situations for using these two devices to enhance the travel experience. The tablet device is mostly used at the hotel (F3, F10, M5). For example M12 uses the tablet device at the hotel e.g. in the morning to plan what to see, while he carries his smart phone while wandering around during the day. The smartphone is smaller and therefore easier to carry around.

"I don't think I need to bring my tablet device on my next trip as we are going to backpack… the smartphone is easier to carry along." (F10)

On the other hand M5 and F3 find the tablet device easier to use due to its screen size. Aging may cause changes in visual capacity and other restrictions for self-arrangements.

"The tablet device is more convenient and sharper, I can see better with it." (M5)

One other hand F14 concluded that she does not need a tablet device as on her iPhone she can e.g. re-size the text.

Many stated that mobile devices are an important information and communication channel during trips nowadays. In fact, mobile devices have to some extent replaced e.g. traditional brochures, paper maps and sending postcards. As M6 stated

"They are so versatile, one can do anything with them; take pictures, search for information, communicate…" (M6)

On the other hand F9 finds it important to have a backup plan e.g. paper map just in case something goes wrong with the mobile device. She feels that she cannot totally rely on her mobile devices. Many of M6s friends find it very nice if he sends them a traditional postcard as no one else is doing that nowadays.

Monetary Burden. Informants said that roaming charges abroad are a problem and depending on the destination they worry about them. For example M12 and F14 have partly guarded themselves from roaming charges in the Nordic and Baltic countries by purchasing a subscription that allows for data transfer at the same price as in Finland. The following three narratives represent well the informants' worries about the potential monetary burden.

"I have to turn off some updates on my phone while travelling abroad." (F10)

"It is better to be careful with the use abroad so that I don't have a big bill waiting for me at home." (F4)

"When I went to the US I had to set the device in flight mode due to terrible roaming-costs." (F3)

Tensions. Both M8 and M12 say that it is very important that the hotel has a proper Wifi, it can even be decisive of whether they will stay at the hotel or not. M12 feels that he will complain if the Wifi is not working properly or if it is over-charged.

"*… I remember once in Sweden they over-charged for the Wifi, then I posted a complaint about that.*" (M12)

M12 brings his mobile devices everywhere, except perhaps to the beach. M12 also admits that he is addicted to his devices.

"I have to admit that I'm addicted …. I feel half naked if I forget my phone." (M12)

Some of them also admit that it can be frustrating to be in network-dead zones, being unable to connect to the Internet. Informants also mentioned that mobile devices make them not really getting away from home (they read what is going in the news at home, read their e-mails, Facebook updates etc.).

"One should turn off the phone…there is not really a need to be online all the time…" (F10)

5 Discussion and Conclusions

First we studied the potential users (five informants) of the Internet with mobile devices during trips. We found a range of possible drivers and barriers, but we sub-divided the results according to four thematic factors influencing the potential use; expected added-value, travel style and partner, knowledge and guidance and anxiety. Generally speaking there was a consensus in this group that there is added-value to use the Internet with mobile devices during a trip. Neither did technology anxiety seem to be a great hurdle, except for one informant. In fact, usage barriers seemed to be more related to their style of traveling, their personal knowledge of using technology and availability of support by e.g. a younger family member. This is in line with previous research that some guidance may be needed to push the adoption and use of mobile device and services by elderly [12]. It may, however, be hard for this group to admit that they need help with technology appliances, which may constitute a greater barrier than being anxious about learning new technologies.

Second we studied the actual users (nine informants) of the Internet with mobile devices during trips. Also in this group we identified a range of possible drivers and barriers, but we sub-divided the results according to five thematic factors influencing the actual use and the travel experience; added-value, travel style, usability, monetary burden and tensions. The findings indicate there is a consensus in the group that there definitely is added-value in using the Internet and mobile devices during trips and it affects the travel experience positively. However, sharing their travel experiences with their mobile devices in social media is not widely used. They also see that they may passivate as travelers and that they may miss out on authentic travel experiences due to

their use of mobile devices. Usability is generally not a problem, rather they reflect about which mobile device(s) to use and bring along. Two of them preferred a tablet device (larger screen) due to restrictions in visual capacity. They all found that a mobile device is handy for their most common style of traveling, self-organized trips. The usage barriers and the negative effects on travel experience are to be found in technology tensions that are caused by e.g. roaming costs in international travel contexts, poor Wifi and dead zones. However, these usage barriers or travel experience barriers are also typical for younger travelers [7, 16]. One of the informants even admitted that he is addicted to his mobile devices.

To sum it up, in this study the conventional description of seniors as technology anxious and highly reluctant to use technology [15, 17] was not widely recognized. Rather this study suggests that there is a substantial number of young-elderly travelers' who are quite advanced in using the Internet with mobile devices during trips. Hence, the study indicates that Internet is nowadays an integral part of many young-elderly travelers' travel experience. However, we should not generalize the results to a total population due to the explorative and qualitative research approach and the purposive sample selection. Quantitative research could empirically verify the results on a larger scale and test dependency between the variables suggested in this study. The analysis of the interview results could also be extended and interpreted to a theoretical model. Limiting the sample to 60–75 year old seniors, here referred to as young-elderly, obviously decreases the importance of health and physical capacity issues to use technology. A 75 + sample may indeed give us very different results [23]. On the other hand, according to Mallenius et al. [12], age is not really the key, rather it is the functional capacity that matters when determining how mobile devices and services are perceived. When studying the behavior of elderly consumers, not only the chronological age should be discussed, but also the cognitive age (how old one "feels" to be) [3, 24]. In this study a substantial number of the informants could not see themselves as seniors by definition.

References

1. Allmér, H., Råberg, M.: Young-elderly and digital use. In: IRIS36 - 36th Information Systems Research Seminar in Scandinavia: "Digital Living", Oslo, Norway (2013)
2. Ajzen, I.: The theory of planned behavior. J. Organ. Behav. Hum. Decis. Process. **50**, 179–211 (1991)
3. Barak, B., Leon, G.S.: Cognitive age: a non chronological age variable. NA – Adv. Consum. Res. **08**, 602–606 (1981)
4. Carlsson, C., Walden, P., Vogel, D., Merne, M.: Young elderly – progressive market for digital service? In: Panel Discussion in the 27th Bled eConference, Bled, Slovenia (2014)
5. Chen, K., Chan, A.H.S.: A review of technology acceptance by older adults. Gerontechnology. **10**(1), 1–12 (2011)
6. Davis, F.D.: Perceived usefulness, perceived ease of use and user acceptance of information technology. J. MIS Q. **13**(3), 319–340 (1989)
7. Eriksson, N.: User categories of mobile travel services. J. Hospitality. Tour. Technol. **5**(1), 17–30 (2014)
8. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research Reading. Addison-Wesley, Massachusetts (1975)

9. Gerpott, T.J.: Attribute perceptions as factors explaining mobile internet acceptance of cellular customers in Germany – an empirical study comparing actual and potential adopters with distinct categories of access appliances. Expert Syst. Appl. **38**(3), 2148–2162 (2011)
10. Hjalager, A.-M., Jensen, J.M.: A typology of travellers based on their propensity to go online before, during and after the trip. In: Fuchs, M., Ricci, F., Cantoni, L. (eds.) Information and Communication Technologies in Tourism 2012, Helsingborg, Sweden, pp. 96–107. Springer, Wien (2012)
11. Douglas, A., Lubbe, B.: Mobile devices as a tourism distribution channel. In: Tussyadiah, I., Xiang, Z. (eds.) Information and Communication Technologies in Tourism 2014, Dublin, Irland, pp. 855–867. Springer, Wien (2014)
12. Mallenius S., Rossi M., Tuunainen V.K.: Factors affecting the adoption and use of mobile devices and services by elderly people -results from a pilot study. In: 6th Annual Global Mobility Roundtable, Los Angeles (2007)
13. McGehee, N.G.: Interview techniques. In: Dwyer, L., Gill, A., Seetaram, N. (eds.) Handbook of Research Methods in Tourism. Edward Elgar, Cheltenham (2012)
14. Minazzi, R., Mauri, A.G.: Mobile technologies effects on travel behaviours and experiences: a preliminary analysis. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, pp. 507–522. Springer, Lugano (2015)
15. Niemelä-Nyrhinen, J.: Baby boom consumers and technology: shooting down stereotypes. J. Consum. Mark. **24**(5), 305–312 (2007)
16. Neuhofer, B., Buhalis, B., Ladkin, A.: Technology as a catalyst of change: enablers and barriers of the tourist experience and their consequences. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, Lugano, Switzerland, pp. 789–802. Springer, Switzerland (2015)
17. Pesonen, J., Komppula, R., Riihinen, A.: Senior travellers as users of online travel. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, Lugano, Switzerland, pp. 831–846. Springer, Switzerland (2015)
18. Paris, C.M., Berger, E.A., Rubin, S., Casson, M.: Disconnected and unplugged: experiences of technology induced anxieties and tensions while traveling. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, Lugano, Switzerland, pp. 803–816. Springer, Switzerland (2015)
19. Reisenwitz, T., Iyer, R., Kuhlmeier, D., Eastman, J.: The elderly´s internet usage: an updated look. J. Consum. Mark. **24**(7), 406–418 (2007)
20. Rogers, E.M.: Diffusion of Innovations, 5th edn. Free Press, New York (2003)
21. Sigala, M.: Measuring customer value in online collaborative trip planning processes. Mark. Intell. Planning **28**(4), 418–443 (2010)
22. Sirakaya, E., Woodside, A.G.: Building and testing theories of decision making by travellers. Tour. Manag. **26**(6), 815–832 (2005)
23. Smith, A.: Older Adults and Technology Use. PewResearchCentre (2014). http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/. Accessed 15 April 2015
24. Sudbury, L., Simcock, P.: A multivariate segmentation model of senior consumers. J. Consum Mark. **26**(4), 251–262 (2009)
25. Statistics Finland: Population projection. http://tilastokeskus.fi/til/vaenn/2012/vaenn_2012_2012-09-28_tie_001_en.html. Accessed 5 March 2015
26. Statistics Finland: Over one-quarter of persons aged 75 to 89 use the Internet. http://www.stat.fi/til/sutivi/2013/sutivi_2013_2013-11-07_tie_001_en.html. Accessed 15 January 2015

27. Thébault, M., Picard, P., Ouedraogo, A.: Seniors and tourism: an international exploratory study on the use of the internet for researching recreational information use. Int. Bus. Res. **6**(3), 22 (2013)
28. Tussyadiah, I.P., Fesenmaier, D.R.: Mediating tourists experiences-access to places via shared videos. Ann. Tour. Res. **36**(1), 24–40 (2009)
29. United Nations: World Population ageing (2013). http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2012_HIGHLIGHTS.pdf. Accessed 5 March 2014
30. Vigolo, V., Confente, I.: Older tourists: an exploratory study on online behavior. In: Tussyadiah, I., Xiang, Z. (eds.) Information and Communication Technologies in Tourism 2014, Dublin, Ireland, pp. 439–449. Springer, Wien (2014)
31. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. J. MIS Q. **27**(3), 425–478 (2003)
32. Venkatesh, V., Tong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. J. MIS Q. **36**(1), 157–178 (2012)
33. Wang, D., Xiang, Z., Fesenmaier, D.R.: Smartphone use in everyday life and travel. J. Travel Res. 1–12 (2014)
34. Xiang, Z., Gretzel, U: Role of social media in online travel information search. Tour. Manag. **31**(2), 179–188 (2010)

An Empirical Study on the Adoption of Online Household e-waste Collection Services in China

Shang Gao[1(✉)], Jinjing Shi[2], Hong Guo[3], Jiawei Kuang[2], and Yibing Xu[2]

[1] Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway
shanggao@idi.ntnu.no
[2] School of Business Administration, Zhongnan University of Economics and Law, Wuhan, China
{ariell236, xyb_mx2013}@163.com, harveykuang.yali@gmail.com
[3] School of Business Administration, Anhui University, Hefei, China
homekuo@gmail.com

Abstract. Online household e-waste collection services are emerging as new solutions to disposing household e-waste in China. This study aims to investigate the adoption of online household e-waste collection services in China. Based on the previous technology diffusion theories (e.g., TAM, UTAUT), a research model with six research hypotheses was proposed in this research. The research model was empirically tested with a sample of 203 users of online household e-waste collection services in China. The results indicated that five of the six research hypotheses were significantly supported. And the most significant determinant for the behavioral intention to use online household e-waste service was effort expectancy. However, facilitating condition did not have significant impact on users' behavior of using online household e-waste collection services.

Keywords: Adoption · UTAUT · Online household · e-waste collection services

1 Introduction

Today, electrical and electronic products become increasingly important in peoples' daily lives. However, it also produces a tremendous amount of e-waste in the mean time. To build a smart and sustainable city, the disposal of waste is a crucial aspect that should be taken into consideration. For example, E-waste has become a serious problem in China in terms of both quantity and toxicity, exacerbated by the development and advancement of electronic industry. A large amount of e-waste was generated in China due to the fast consumption rates of electrical and electronic products. E-waste was largely collected by the informal sector. According to the previous study (e.g., [22]), it was found that the majority of electrical and electronic

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 36–47, 2015.
DOI: 10.1007/978-3-319-25013-7_4

products were often sold to peddlers. However, most peddlers tended to focus on waste fractions of the collected e-waste with which they can make a profit instead of focusing on raised environmental problems.

With the rapid economic growth in China, consumers' demand for electrical and electronic products is booming simultaneously. We have witnessed exponential increase of the quantity of output of four major household appliances (e.g., TV) in the past two decades. For instance, the output of TVs has increased to 133.82 million units in 2014, from 33.70 million units in 1998. When these household appliances eventually reach the ends of their life circles, they will become obsolete and pose a potential threat to both natural environment and human health if they were not properly disposed. The recycling chain for e-waste is classified into three main subsequent steps: collection, sorting/dismantling and pre-processing (including sorting, dismantling and mechanical treatment) and end-processing [20]. Being the first step of the whole recycling chain, collection serves as a crucial process of e-waste recycling. Currently, there exist six major options for household e-waste disposal: second use, storage, discarding, disassembly, formal collection and informal collection [6]. Online household e-waste collecting and recycling platform is aiming to become a convenient way of collecting e-waste and to offer higher recycling price than official price notwithstanding. However, it seems that most Chinese electrical and electronic products consumers are not familiar with this platform.

Online household e-waste collection revolutionizes the traditional ways of collecting and disposing e-waste. It integrates information flows, logistics and capital flows, constructing a closer connection between e-waste sellers and recyclers, which facilitates the e-waste collecting process and ultimately improves the effectiveness and efficiency of the whole recycling chain for e-waste.

The objective of this research is to investigate consumers' behavior intention to use online household e-waste collection services in China. Based on previous technology diffusion theories (e.g., TAM [7], UTAUT [26]), a research model with six research hypotheses was proposed. And the research model was empirically tested using data collected from a survey of users of an online household e-waste collection service called Taolv365 in China.

The remainder of this paper is organized as follows. The related literature review is reviewed in Sect. 2. The research model and hypotheses are presented in Sect. 3. In Sect. 4, an empirical study is carried out to examine the research model. This is followed by a discussion of the findings and limitation of the study in Sect. 5. Section 6 concludes this research work and points out directions for future research.

2 Literature Review

2.1 Technology Diffusion Theories

An important and long-standing research question in information systems research is how to accurately explain user adoption of information systems [8]. Several models have been developed to test the users' attitude and intention to adopt new technologies or information systems. These models include the Technology Acceptance Model

(TAM) [7], Theory of Planned Behavior (TPB) [1], Innovation Diffusion Theory (IDT) [19], Unified Theory of Acceptance and Use of Technology (UTAUT) [26], and Mobile Services Acceptance Model (MSAM) [11, 12]. UTAUT was developed through a review and consolidation of the constructs of eight models (e.g., TAM, TRA [9], TPB, IDT) that earlier research had employed to explain IS usage behavior. Moreover, variables that influence users' behavioral intention employed in TAM3 [23] were classified into the following four types: individual differences, system characteristics, social influence and facilitating conditions [24].

Due to its unified perspective, although UTAUT has not been as widely used as TAM, partly resulting from its complexity and intricacy to test its applicability, it has gradually drawn researchers' attentions and has been recently applied to exploring user acceptance of mobile technologies and mobile devices (e.g., [4, 14, 17]). We built our research model based on UTAUT to assess the adoption of online household e-waste collection services in China.

2.2 Research on e-waste Collection Services in China

As the largest exporter of electrical and electronic products and importer of waste electrical and electronic products around the world, China plays a key role in the global life cycle of electronics [28]. Being concerned about the dispose of e-waste, many studies have been conducted during the last few decades.

As shown in Table 1, previous studies have provided us with different aspects to understand the current situations of e-waste collection and recycling system in China, including the legislative policy, e-waste collection channels, household recycling behaviors, etc. However, they were mainly focusing on the traditional methods of e-waste dispose while neglecting online e-waste collection services.

This research aims to complement and extend existing studies on the adoption of e-waste collection services by focusing on the emerging channel of disposing e-waste: online household e-waste collection services. The online household e-waste collection services can be seen as information systems. Therefore, we would like to use the existing technology diffusion theories to examine the adoption of online household e-waste collection services in China.

3 Research Model and Hypotheses

A research model that identifies important factors that impact users' intention to use online household e-waste collection services was developed in this research. The proposed research model (see Fig. 1) is a simplified version of UTAUT. We have developed the following six research hypotheses (labeled in Fig. 1) based on the research model.

Performance Expectancy and Effort Expectancy. Performance expectancy is defined as the degree to which an individual believes that using the system will help him or her to attain gains in job performance [26]. It reflects user perception of performance improvement by using online household e-waste collection services such as convenience, fast response, and service effectiveness. The performance expectancy

Table 1. Literature review on e-waste collection system in China

Literature	Research purpose	Findings
Streicher-Porte, Geering, 2009 [22]	Analyze household data and compares literature sources about how Chinese households dispose of obsolete electrical and electronic equipment.	The findings indicated that the informal collection sector of obsolete EEE in China mainly in terms of door-to-door collection by peddlers played a major role in the current management of this waste stream.
Yu, Williams, Ju, & Shao, 2010 [28]	Review the existing framework for e-waste management in China including regulatory policies and pilot projects.	Two alternate policies were proposed: shared responsibility with deposit to incentivize consumer participation and integrating informal collection/reuse with formal dismantling/recycling.
Chi, Streicher-Porte, Wang, & Reuter, 2011 [5]	Gather information on informal e-waste management, take a look at its particular manifestations in China	The findings revealed the actual situation of e-waste management in China. Moreover, it indicated that the improvement of informal recycling sector lied on dedicated efforts from economic, technical and social aspects.
Chi, Wang, & Reuter, 2014 [6]	Investigate the collection channels of e-waste and household recycling behaviors in Taizhou city of China.	The authors suggested that rather than directly competing with the informal collection sector, a better solution was to harness its strengths and incorporate it into a more accountable and regulated e-waste collection system.

construct is the strongest predictor of behavioral intention and remains significant at all settings [26]. When users feel that online household e-waste collection services are useful and can bring them convenience to dispose their e-waste, their intention to use to the services are likely to be high.

Effort expectancy is defined as the degree of ease associated with the use of the system [26]. It reflects user perception of how difficult it is to use online household e-waste collection services. According to earlier research (e.g., [26]), effort expectancy positively affects performance expectancy. When users feel that online household e-waste collection services are easy to use and do not require much effort, they will have a high expectation toward acquiring the expected performance. Otherwise, their

performance expectancy is likely to be low. Thus, we formulate the following research hypotheses:

H1: Effort expectancy has a positive influence on performance expectancy.
H2: Performance expectancy has a positive influence on users' behavioral intention of using online household e-waste collection services.
H3: Effort expectancy has a positive influence on users' behavioral intention of using online household e-waste collection services.

Social Influence. Social influence is similar to subjective norm of TRA and reflects the effect of environmental factors such as the opinions of a user's friends, relatives, and superiors on user behavior [15]. In addition, recommendation as well as word-of-mouth effect will also affect consumers' behavioral intention [10]. Thus, we formulate the following research hypothesis:

H4: Social influence has a positive influence on users' behavioral intention of using online household e-waste collection services.

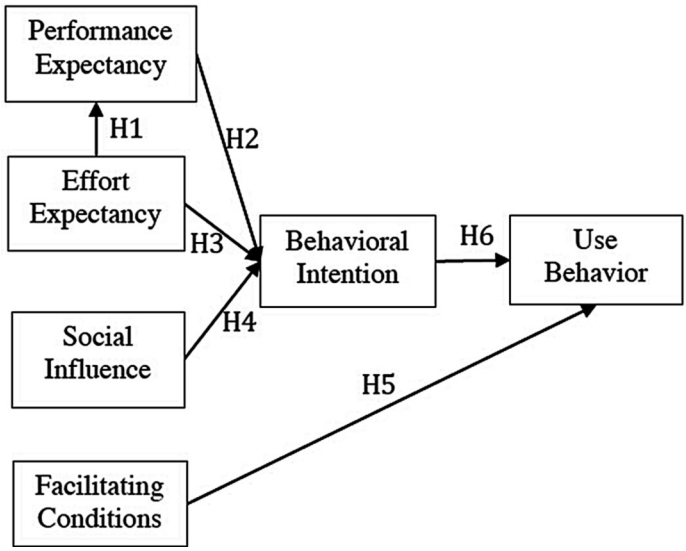


Fig. 1. Research model

Facilitating Conditions. Facilitating conditions are similar to perceived behavioral control of TPB and reflect the effect of a certain user's knowledge, ability, and resources [26]. Online household e-waste service is a rather new service for e-waste recycling, but similar to other e-commerce services. In order to use this service, users need some basic infrastructures like PC, smart phones, Internet, etc. In addition, users need to bear usage costs such as data service and logistics fees when using online household e-waste collection services. If users do not have these necessary devices as

well as financial resources and operational skills, they are not likely to adopt or use online household e-waste collection services. Thus, we formulate the following research hypothesis:

H5: Facilitating conditions have a positive influence on users' behavior of using online household e-waste collection services.

Behavior Intention. Consistent with the underlying theory for all of the intention models [21], we expect that behavioral intention will have a significant positive influence on usage behavior [26]. If the intention of a certain user to use online household e-waste collection services is high, he or she will be more likely to use this service. Otherwise, the user is unlikely to adopt this service. Thus, we formulate the following research hypothesis:

H6: Behavioral intention has a positive influence on use behavior.

4 An Empirical Study with the Research Model

In this empirical test, our research model was examined through the use online household e-waste collection service called Taolv365.

4.1 The Online Household e-waste Collection Service: Taolv365

Taolv365 aimed to build online household recycling system with Chinese characteristics. It is one of the first pioneers using the Internet as the basis for the reverse supply service platform. In addition, it has largest amount of registered users in China. Therefore, this website is a good case for instrument testing and identifying relevant implications for practitioners.

Based on the national environmental standards, Taolv365 gradually establishes the system of online household e-waste collection services to manage the chaotic e-waste collection market. Taolv365 also aimed to deal with some existing recycling problems (e.g., the lack of channels). Their ambition is to create an intelligent and green recycling system to collect e-waste in a more convenient and environmental-friendly way.

Taolv365 provides the following three services to the users:

- Information Service: Taolv365 can provide information to the clients in the following channels: call center, online messaging service, and the third platform collecting channels which can offer 7days/24 h help to users. Anyone who has e-waste can choose one of the channels for help.
- Price Inquiry Service: E-waste is classified on Taolv365. Users can easily get the price of their e-waste according to the standard classification system. It also provides SMS platform for immediate feedback, professional customer service so that users can timely grasp of recycling market and trends. Users can also donate their devices or electronics to charities via Taolv365.

- The Transaction Channel Service: The website integrates recycling industry buyers through formal channels. The buyers bid online, commit the transaction volume and shipments, and classify the standards carefully so as to achieve the optimal price. As long as submitting the order, the service can get the real-time tracking of the logistics, trading and payment situations.

4.2 Instrument Development

The validated instrument measures from previous research were used as the foundation to create the instrument for this study. Previous studies were reviewed to ensure that a comprehensive list of measures were included. In order to ensure that the instrument better fit this empirical study, some minor words changes were made to ensure easy interpretation and comprehension of the questions. All the items were adopted from prior studies [26] and modified to fit the domain of online household e-waste collection services. As a result, 17 measurement items[1] were included in the questionnaires. In addition, a seven-point Likert scale, with 1 being the negative end of the scale (strongly disagree) and 7 being the positive end of the scale (strongly agree), was used to examine participants' responses to all items in the survey.

4.3 Samples

The data for this study was collected through paper-based questionnaires from 20th April to 30th April 2015 in the biggest city in the central China. People were asked to participate in the survey voluntarily. Firstly, we explained who we were, what they were supposed to do during our survey, and the purpose of the survey. The participants were also informed that the results would be reported only in aggregate and their anonymity would be assured. After participants experienced the online household e-waste collection service provided by Taolv365, they were asked to complete the questionnaires and submitted them to us. 210 completed questionnaires were collected, among which 203 of them were valid questionnaires (i.e., valid respondent rate 96.7 %). Among the participants, 86 of the participants were male, and 117 were female. In terms of age, 120 participants were 25 years old and under 25 years old, while 83 participants were over 25 years old.

4.4 Descriptive Results

Some key findings from the descriptive results are summarized in this section. The means for each of the items in the survey are presented in Table 2. The mean value for the measurement value PE1 is 5.30. 145 of the 203 respondents (71.43 %) agreed that the service of online household e-waste collection was useful. In the same time, 144 respondents felt that online household e-waste collection service improved both their efficiency and convenience of handling e-waste.

[1] The survey items are available at this link: http://www.idi.ntnu.no/~shanggao/ewaste.html.

According to the results, most respondents thought that it would be easy for them to use online household e-waste collection services. 132 respondents indicated that it is easy for them to learn how to use online household e-waste collection services. Moreover, gender differences had an obvious effect on peoples' reaction towards social influence. Men were more likely to be affected by someone who were important to or had influence on them than women. Similarly, effort expectancy has a more significant likelihood to result in peoples' behavioral intention among men than women.

Item FC1 has the highest mean value in the construct facilitating conditions. Most respondents indicated that they had the necessary resources to use online household e-waste collection services. Furthermore, many of them thought that they had the necessary knowledge to use online household e-waste collection services.

Furthermore, many respondents gave us a neutral response when we were asking the question on their frequency to use online household e-waste collection services to deal with their household appliances or electronic devices. This means that there is still a huge business market for online household e-waste collection services in China.

Table 2. Means, factor loadings, composite reliability, and AVE for each item

Construct	Item	Mean	Factor loading	Composite reliability	AVE	Cronbach's alpha
Performance expectancy	PE1	5.30	0.850	0.927	0.810	0.882
	PE2	5.16	0.925			
	PE3	5.26	0.923			
Effort expectancy	EE1	4.68	0.932	0.950	0.827	0.930
	EE2	4.58	0.919			
	EE3	4.98	0.898			
	EE4	4.65	0.888			
Social influence	SI1	4.51	0.965	0.968	0.937	0.933
	SI2	4.47	0.917			
Facilitating conditions	FC1	5.17	0.867	0.913	0.778	0.867
	FC2	5.15	0.877			
	FC3	4.70	0.901			
Behavioral intention	BI1	4.88	0.939	0.946	0.854	
	BI2	4.86	0.918			0.914
	BI3	4.72	0.914			
Use behavior	UB1	4.02	0.974	0.973	0.947	0.944
	UB2	4.11	0.973			

4.5 Measurement Model

The quality of the measurement model is determined by (1). Content validity, (2). Construct reliability and (3). Discriminant validity [2]. To ensure the content validity of our constructs, a pretest with 6 researchers in the field of information systems was carried out in March 2015. And we found that the questionnaire was well understood by all the researchers.

Table 3. Discriminant validity

Variables	BI	EE	FC	PE	SI	UB
BI	0.924					
EE	0.755	0.909				
FC	0.825	0.803	0.882			
PE	0.730	0.732	0.713	0.900		
SI	0.760	0.714	0.667	0.649	0.968	
UB	0.670	0.593	0.538	0.521	0.693	0.973

Note: Diagonals represent the average variance extracted, while the other matrix entries represent the squared correlations.

To further test the reliability and validity of each construct in the research model, the Internal Consistency of Reliability (ICR) of each construct was tested with Cronbach's Alpha coefficient. As a result, the Cronbach's Alpha values range from 0.867 to 0.944. A score of 0.7 is marked as an acceptable reliability coefficient for Cronbach's Alpha [18]. All the constructs in the research model were above 0.70. Consequently, the scales were deemed acceptable to continue.

Convergent validity was assessed through composite reliability (CR) and the average variance extracted (AVE). Bagozzi and Yi [3] proposed the following three measurement criteria: factor loadings for all items should exceed 0.5, the CR should exceed 0.7, and the AVE of each construct should exceed 0.5. As shown in Table 3, all constructs were in acceptable ranges.

4.6 Structural Model and Hypotheses Testing

The structural model was tested using SmartPLS. Table 4 presents the path coefficients, which are standardized regression coefficients. Five (H1, H2, H3, H4, H6) of the six research hypotheses were significantly supported. According to the results, performance expectancy, effort expectancy and social influence were found to have a statistically significant effect on users' behavior intention to use online household e-waste collection services, while facilitating condition did not have significant impact on users' behavior of using online household e-waste collection services.

Table 4. Test of hypotheses based on path coefficient

Hypothesis	Path coefficient	Hypothesis result
H1.Effort expectancy to performance expectancy	0.732***	Supported
H2.Performance expectancy to behavior intention	0.279**	Supported
H3.Effort expectancy to behavior intention	0.281**	Supported
H4.Social influence to behavior intention	0.379**	Supported
H5.Facilitating conditions to behavior intention	–0.044	Not supported
H6.Behavior intention to use behavior	0.706***	Supported

*$p < 0.05$; **$p < 0.01$; *** $p < 0.001$

The R^2 (R square) denotes to coefficient of determination. It provides a measure of how well future outcomes are likely to be predicted by the model, the amount of variability of a given construct. In our analysis, the R^2 coefficient of determination is a statistical measure of how well the regression coefficients approximate the real data point. According to the result, 70.4 % of the variance of behavior intention can be explained by the research model.

5 Discussion

In this research, we studied the adoption of online household e-waste collection services in China. From the theoretical perspective, this research contributed to the literature on the adoption of new technologies (i.e., online e-waste collection services) in China by building upon previous technology diffusion theories. To our best knowledge, most previous studies were mainly focusing on the traditional channels of household collection as well as the conflicts between formal and informal e-waste recycling systems (e.g., [6, 22]). This research is one of the first studies to examine users' adoption of online household e-waste collection services in China. From a practical perspective, it offered some insights to the use and adoption of online household e-waste collection services in China.

Since effort expectancy has an indirect effect on users' behavioral intention through performance expectancy, the total effect of effort expectancy on behavioral intention is 0.281 + 0.732 * 0.279 = 0.485. Therefore, the most important determinant for the behavioral intention to use online household e-waste service was effort expectancy, while social influence and performance expectancy were ranked as second and third respectively. As to the significant impact of social influence on users' behavior intention, one possible explanation is that Chinese consumers are considered to be more concerned about their social images. Moreover, the effect of word-of-mouth is an important influencer in China [27]. In addition, effort expectancy has an significant positive influence on performance expectancy, which is consistent with previous research findings [26]. However, facilitating conditions is not proved to be the determinant for the use behavior. This can be partially attributed to the fact that most participants in this study were between 15 to 25 years old. They were quite skillful at using online services and unlikely to ask for help from their fellows. This had been further confirmed in the follow-up interview session. Some participants also indicated that online household e-waste collection services were easy to use so that they did not need help from others. Thus, facilitating conditions did not have a significant positive influence on use behavior. This is also consistent with the previous findings (e.g., [13, 25]).

When promoting the online household e-waste collection services, practitioners could make advantage of social influence, such as celebrity worship and the effect of word-of-mouth, especially in countries like China. Moreover, most participants believed that online household e-waste collection services improve the efficiency to deal with household e-waste. It was also interesting to note that people with different gender and age groups responded differently to social influence. The results indicated that men were much more likely to be affected by someone who were important or have influence on them than women.

However, we were also aware of some limitations. Firstly, we only tested the research model and research hypotheses with samples from one province in China. This sample might not be fully representative of the entire population in China. Secondly, all the data were collected using self-reported scales in the research. This may lead to some caution because common method variance may account for some of the results that has been cited as one of the stronger criticisms of tests of theories with TAM and TAM-extended research [16]. However, our data analysis with convergent and discriminant validity does not support the presence of a strong common methods bias. Last but not least, the services provided by Taolv365 may not represent all the online household e-waste collection services.

6 Conclusion and Future Research

This research was designed to explore users' adoption of online household e-waste collection services in China. A research model with six research hypotheses was proposed. The results indicated that five of the six research hypotheses were significantly supported. And the most significant determinant for the behavioral intention to use online household e-waste service was effort expectancy. However, facilitating condition did not have a significant positive impact on users' behavior of using online household e-waste collection services.

Continuing with this stream of research, we plan to examine some additional constructs' (e.g., trust) influence on the adoption of online household e-waste collection services. Future research is also needed to empirically verify the research model with larger samples across China.

References

1. Ajzen, I.: The theory of planned behavior. Organ. Behav. Hum. Decis. Process. **50**(2), 179–211 (1991)
2. Bagozzi, R.P.: The role of measurement in theory construction and hypothesis testing: toward a holistic model. In: Ferrell, O.C., Brown, S.W., Lamb, C.W. (eds.) Conceptual and Theoretical Developments in Marketing, pp. 15–32 (1979)
3. Bagozzi, R.P., Yi, Y.: Specification, evaluation, and interpretation of structural equation models. J. Acad. Mark. Sci. **40**(1), 8–34 (2012)
4. Carlsson, C., Carlsson, J., Hyvonen, K., et al.: Adoption of Mobile Devices/Services - Searching for Answers with the UTAUT Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 06. IEEE Computer Society (2006)
5. Chi, X., Streicher-Porte, M., Wang, M.Y., et al.: Informal electronic waste recycling: a sector review with special focus on China. Waste Manag. **31**(4), 731–742 (2011)
6. Chi, X., Wang, M.Y., Reuter, M.A.: E-waste collection channels and household recycling behaviors in Taizhou of China. J. Clean. Prod. **80**, 87–95 (2014)
7. Davis, F.D.: Perceived usefulness, perceived ease of use and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
8. DeLone, W., McLean, E.: Information systems success: the quest for the dependent variable. Inf. Syst. Res. **3**(1), 60–95 (1992)

9. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975)
10. Fitzsimons, G.J., Lehmann, D.R.: Reactance to recommendations: When unsolicited advice yields contrary responses. Mark. Sci. **23**(1), 82–94 (2004)
11. Gao, S., Krogstie, J., Gransæther, P.A.: Mobile Services Acceptance Model. In: The Proceedings of International Conference on Convergence and Hybrid Information Technology. IEEE Computer Society (2008)
12. Gao, S., Krogstie, J., Siau, K.: Adoption of mobile information services: an empirical study. Mob. Inf. Syst. **10**(2), 147–171 (2014)
13. Gao, S., Krogstie, J., Yang, Y.: Differences in the adoption of smartphones between middle aged adults and older adults in China. In: Zhou, J., Salvendy, G. (eds.) ITAP 2015. LNCS, vol. 9193, pp. 451–462. Springer, Heidelberg (2015)
14. Gao, S., Yang, Y., Krogstie, J.: The adoption of smartphones among older adults in china. In: Liu, K., Nakata, K., Li, W., Galarreta, D. (eds.) ICISO 2015. IFIP AICT, vol. 449, pp. 112–122. Springer, Heidelberg (2015)
15. López-Nicolás, C., Molina-Castillo, F.J., Bouwman, H.: An assessment of advanced mobile services acceptance: contributions from TAM and diffusion theory models. Inf. Manag. **45** (6), 359–364 (2008)
16. Malhotra, N.K., Kim, S.S., Patil, A.: Common method variance in IS research: a comparison of alternative approaches and a reanalysis of past research. Manage. Sci. **52**(12), 1865–1883 (2006)
17. Park, J., Yang, S., Lehto, X.: Adoption of mobile technologies for Chinese consumers. J. Electron. Commer. Res. **8**(3), 196–206 (2007)
18. Robinson, J.P., Shaver, P.R., Wrightsman, L.S.: Criteria for scale selections and evaluation. Academic Press, San Diego (1991)
19. Rogers, E.M.: The diffusion of innovations. Free Press, New York (1995)
20. Schluep, M., Hagelueken, C., Kuehr, R., et al.: Sustainable Innovation and Technology Transfer Industrial Sector Studies: Recycling–from E-waste to Resources. United Nations Environment Programme and United Nations University, Bonn (2009)
21. Sheppard, B.H., Hartwick, J., Warshaw, P.R.: The theory of reasoned action: a meta-analysis of past research with recommendations for modifications and future research. J. Consum. Res. **15**(3), 325–343 (1988)
22. Streicher-Porte, M., Geering, A.-C.: Opportunities and threats of current e-waste collection system in China: a case study from Taizhou with a focus on refrigerators, washing machines, and televisions. Environ. Eng. Sci. **27**(1), 29–36 (2010)
23. Venkatesh, V., Bala, H.: TAM 3: advancing the technology acceptance model with a focus on interventions. Manuscript in Preparation (2013). http://www.vvenkatesh.com/IT/organizations/Theoretical_Models.asp
24. Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. Decis. Sci. **39**(2), 273–315 (2008)
25. Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag. Sci. **46**(2), 186–204 (2000)
26. Venkatesh, V., Morris, M.G., Davis, G.B., et al.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
27. Wiedemann, D.G., Haunstetter, T., Pousttchi, K.: Analyzing the basic elements of mobile viral marketing-an empirical study. In: 7th International Conference on Mobile Business, ICMB 2008, pp. 75–85. IEEE (2008)
28. Yu, J., Williams, E., et al.: Managing e-waste in China: policies, pilot projects and alternative approaches. Resour. Conserv. Recycl. **54**(11), 991–999 (2010)

Perceptions of Teachers and Guardians on the Electronic Record in the School-Family Communication

António Abreu[1], Álvaro Rocha[2(✉)], and Manuel Pérez Cota[3]

[1] ISCAP-IPP, Rua Jaime Lopes Amorim,
s/n 4465-004 S. Mamede de Infesta,
Matosinhos, Portugal
aabreu@iscap.ipp.pt
[2] DEI, University of Coimbra,
Pólo II – Pinhal de Marrocos,
3030-290 Coimbra, Portugal
amrocha@dei.uc.pt
[3] EEI, University of Vigo, Campus Torrecedeira,
Rúa Torrecedeira 86, 36280 Vigo, Spain
mpcota@uvigo.es

Abstract. This paper presents the first stage of an investigation work whose purpose is to introduce a new form of communication between the school and the family, through an electronic record (ER). We tried to identify, with a questionnaire, the perceptions of Teachers and Guardians towards the adoption of an ER. Both groups underlined the importance of communication between the school and the family, and expressed interest in the ER, as a communicational tool.

Keywords: School · Family · School-family relationship · Communication · Electronic record

1 Introduction

The present study, focused on the communication, management and dissemination of information in digital environments, is of great significance and interest. We intent to give our contribution to the communication between the school and the family, with the development of an Electronic Record (ER).

Indeed, the technological revolution imposed changes both in society [1, 11, 12, 16, 17, 28] and in the forms of communication, which exceed barriers of space and time [6, 7, 22].

Schools should use electronic resources to develop new forms of communication between the school and the family [3, 4, 17, 23–25, 28], hence de ER [12–15, 18, 29].

Currently, this communication is established via the traditional student record (TR), in paper format, which is not the most direct contact vehicle, as it requires the mediation of the student and, moreover, limits the communication content [18].

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 48–62, 2015.
DOI: 10.1007/978-3-319-25013-7_5

In order to render this communication easier [17], the need make the TR a more dynamic and attractive tool, both to families and the school, emerged, and this may be achieved with the ER.

We intend to develop a study that assumes as a starting point three investigation questions:

- Is the school-family communication fundamental for the teaching-learning process?
- Is the TR a fundamental tool for school-family communication?
- Does the ER improve and increase the potential of school-family communication?

This study comprised three stages. The first stage, the focus of this paper, involved a feasibility study as to the adoption of the ER in school-family relationships, based on the perceptions of Guardians and Teachers.

2 Investigation Methodology

In this stage of the study, we adopted the survey investigation method, to the extent that this constitutes a viable and desirable method within a school environment and generates accurate and reliable measures which allow for a statistical analysis [6]. The adopted tool was the questionnaire, divided in two parts: the first focused on the characterization of the sample and the latter being the questionnaire itself, which, in its turn, included Part B, pertaining to the "Student Record", and Part C, concerning the "Inclusion of items in an Electronic Report". We adopted a *Likert* type ordinal scale with five answer alternatives (from "1" to "5") that ranged from "Completely disagree" to "Completely agree".

For this investigation work we developed the following hypothesis:

H1: The educational community values communication between the school and the family;
H2: The TR is a fundamental tool in school-family communication;
H3: The ER improves school-family communication.

We conducted a pre-test in a group of thirty Teachers and Guardians, in order to clarify and validate the questionnaire. As the questionnaire did not offer doubts, we moved to the questionnaire distribution stage, in paper format and on-line, amongst the participating schools and school groupings, considering the answering availability on the part of the respondents and according to the preferences of School Principals.

The survey was carried in the form of a questionnaire in schools and cooperative/private and public schools located in the North, Centre and South of the country, between 17 November 2013 and 8 March 2014. In an informal meeting the School Principals were informed as to the purpose of the study and the entire questionnaire administration process, and they in turn, together with the responsible parties of the school groupings, had to explain Teachers, class directors, elements from parent associations and Guardians the purpose of the investigation work and how they could participate.

The obtained data was treated with the SPSS computer program, version 20. Besides the descriptive data analysis, we carried out a correlational analysis, crossing

some of the variables comprised by the questionnaire, in order to find the answers to the three laid out investigation hypothesis [19]. We also used absolute (standard deviation) and relative (coefficient of variation) measures of dispersion, useful when describing the observed variation in the values of a set and informing us as to the homogeneity. The smaller (<1) the standard deviation and (<=20 %) the coefficient of variation, the smaller the dispersion and, consequently, the more homogenous the answers.

2.1 Characterization of the Sample

The sample included 1002 Guardians of students attending basic and secondary education and 300 Teachers exercising their professional activity in cooperative/private and public educational institutions.

2.2 Factor Analysis

We carried out a data analysis to the questions comprising Part B and Part C of the Guardians and Teachers questionnaires, adopting data treatment techniques that allowed us to agglomerate the initial information, with a view to render the analysis easier [20].

We selected the principal components method, in order to isolate the original correlating variables, constituting factors [21].

The factor analysis method analyses a set of variables with the purpose of verifying the possibility of grouping answers that are similarly interpreted by the elements of the sample, in order to determine their position within the set of variables [9, 26, 27]. This allows us to turn the variables of a scale into a smaller number of factors: the principal components. To define the number of components to be retained, we carried out a preliminary analysis, applying the factor analysis and interpreting the obtained results [8].

In part B of the Teachers and Guardians questionnaires, concerning the "Student Record", we drew our factors from the 15 variables presented in Table 1.

We began our analysis with the de *Kaiser Meyer Olkin (KMO)* and the *Bartlett* tests, which determine the quality of the factor analysis to the sample data [2, 10].

After analysing the test results, the *KMO* revealed a value of 0.887 and the *Bartlett* a significance value below 5 %, the factor analysis to component B of the Guardians and Teachers questionnaires was considered feasible [11].

Using the principal components method, we proceeded with the extraction of factors from the 15 variables, adopting the *Kaiser* criterion (eigenvalues above one), complying with the requirement that factors should explain at least approximately 60 % of the total variation observed in the original variables [10]. We retained four factors, which explain 66.7 % of the total variation observed in the original 15 variables.

Through the *Varimax* method we were able to obtain the extreme values of the coefficients relating each variable to the retained factors, in order to associate each variable with a single factor [5]: Factor 1 – ER; Factor 2 – Impact of the school-family relationship in the teaching learning process; Factor 3 – TR; Factor 4 – School-family relationship difficulties.

Table 1. Part B of the Teachers and Guardians questionnaires - "Student Report".

Part B of the Teachers and Guardians questionnaires – "Student Record".	
#	Item
1	The relationship between the family and the school is fundamental for the teaching-learning process
2	The relationship between the school and the family is not always easy and stands out as a complex relationship
3	Whenever the parents are involved the development of the students is potentially improved
4	The participation of parents in the school benefits the Teachers and contributes to a more constructive work on their part
5	Parents assume more favourable attitudes towards Teachers when they cooperate constructively with each other
6	Parents can easily head to the school during the receiving hours appointed by Class Directors and in meeting days
7	The Traditional Student Record constitutes an important tool in family-school communications
8	Teachers regularly use the Traditional Record in their communications with Guardians
9	It is important to find a tool that will allow for the intensification of and an easier communication between the School and the Guardian
10	The adoption of an Electronic Report may contribute to an improved communication between the school and the Guardian
11	An Electronic Record may constitute an important work tool for Class Directors, Teachers, Parents and School Principals
12	An Electronic Report may constitute a communication resource between every Teacher and Guardians
13	An Electronic Report may be used by School Principals to contact Teachers
14	An Electronic Report may be used by School Principals to contact Guardians
15	An Electronic Record offers more advantages than the Traditional Record where the communication between the multiple school community actors is concerned

Table 2 presents the factorial matrix after the *Varimax* rotation, where the saturation between each factor and the principal components can be observed. The saturations of the variables in each factor are always higher than the required minimum of 40 %. The value presented in Table 2 – 0,661, is owed to the fact that the respective item is in a reverse scale. In schools, at the beginning of the school year, the Class Director defines, according to his school hours, a day and a time to receive Guardians. These receiving hours usually coincide with the opening hours of the educational establishment. However, for most of the Guardians this schedule overlaps their office hours. To make this situation worse, many Guardians work outside their areas of residence, where educational establishments are usually located, which prevents them from attending the pre-scheduled meetings. Being a negative value, the inclusion of item 6 in factor 4 is carried out in a reverse scale.

In part C of the questionnaires ("Inclusion of items in an Electronic Report"), Teachers and Guardians were inquired about 7 items, as seen in Table 3.

The *Kaiser Meyer Olkin (KMO)* and *Bartlett* tests revealed a *KMO* value of 0.881 and a Bartlett significance value below 5 %, which allowed us to carry out the factor analysis, as the results were considered significant [2, 10].

After confirming the possibility of carrying out a factor analysis, and adopting the method of the principal components, we obtained our factors from the 7 variables, employing the *Kaiser* criterion (eigenvalues above one) and complying with the criterion that factors should explain at least approximately 60 % of the total variation observed in the original variables [10]. We retained two factors, which explain 72.6 % of the total variation observed in the 7 original variables [27].

Table 2. Rotation method: Varimax with Kaiser normalization. Rotation converged in 5 iterations. N = 1302. KMO = 0.887.

Items	Factor			
	1	2	3	4
1. The relationship between the family and the school is fundamental for the teaching-learning process	0.050	**0.702**	0.017	–0.057
2. The relationship between the school and the family is not always easy and stands out as a complex relationship	0.065	0.067	0.092	**0.830**
3. Whenever the parents are involved the development of the students is potentially improved	0.124	**0.741**	0.084	0.125
4. The participation of parents in the school benefits the Teachers and contributes to a more constructive work on their part	0.166	**0.724**	0.104	–0.023
5. Parents assume more favourable attitudes towards Teachers when they cooperate constructively with each other	0.101	**0.749**	0.117	–0.052
6. Parents can easily head to the school during the receiving hours appointed by Class Directors and in meeting days	–0.095	0.155	0.295	**–0.661**
7. The Traditional Student Record constitutes an important tool in family-school communications	0.024	0.220	**0.808**	–0.074
8. Teachers regularly use the Traditional Record in their communications with Guardians	0.052	0.042	**0.873**	–0.018
9. It is important to find a tool that will allow for the intensification of and an easier communication between the School and the Guardian	**0.474**	0.286	0.100	0.341
10. The adoption of an Electronic Report may contribute to an improved communication between the school and the Guardian	**0.886**	0.086	–0.019	0.109

(*Continued*)

Table 2. (*Continued*)

Items	Factor			
	1	2	3	4
11. An Electronic Record may constitute an important work tool for Class Directors, Teachers, Parents and School Principals	**0.899**	0.137	–0.011	0.074
12. An Electronic Report may constitute a communication resource between every Teacher and Guardians	**0.904**	0.118	–0.021	0.074
13. An Electronic Report may be used by School Principals to contact Teachers	**0.698**	0.146	0.147	–0.014
14. An Electronic Report may be used by School Principals to contact Guardians	**0.896**	0.080	0.045	0.017
15. An Electronic Record offers more advantages than the Traditional Record where the communication between the multiple school community actors is concerned	**0.880**	0.038	–0.033	0.052
% Explained variance	31.8	15.7	10.5	8.6
Total explained variance	66.7			

Table 3. Part C of the Teachers and Guardians questionnaires – "Inclusion of items in an Electronic Report"

Part C of the Teachers and Guardians questionnaires - "Inclusion of items in an Electronic Report"	
Number	Item
1	Absence justifications by the Guardians
2	Absence management by the Class Director
3	Mid-term evaluation record
4	Evaluation criteria and tools for each subject
5	Student self-assessment forms
6	Communication management (Messages, *SMS* and *Emails*) between Class Directors, Teachers and Guardians
7	Videoconference between Class Directors and Guardians

Through the *Varimax* method we were able to obtain the extreme value of the coefficients relating each variable to the retained factors, connecting each variable to a single factor [5]: Factor 1 – evaluation and absence records, Factor 2 – student self-assessment and communication.

Table 4 shows the factorial matrix after the *Varimax* rotation, where the saturations between each factor as well as the principal components can be observed. The saturation of the variables in each factor stand above the required minimum of 40 %.

Table 4. Rotation Method - Varimax with Kaiser normalization. Rotation converged in 5 iterations. N = 1302. KMO = 0,887.

Items	Factor	
	1	2
1. Absence justifications by the Guardians	**0.861**	0.141
2. Absence management by the Class Director	**0.868**	0.248
3. Mid-term evaluation record	**0.767**	**0.405**
4. Evaluation criteria and tools for each subject	**0.588**	**0.571**
5. Student self-assessment forms	**0.439**	**0.667**
6. Communication management (Messages, *SMS* and *Emails*) between Class Directors, Teachers and Guardians	**0.458**	**0.668**
7. Videoconference between Class Directors and Guardians	0.054	**0.884**
% Explained variance	40.5	32.1
Total explained variance	72.6	

In Part C of the questionnaires – "Inclusion of Items in an Electronic Report", the inclusion of two factors was substantiated. However, several items converged in more than one, and this justifies a subsequent analysis of the items individually (Table 4).

2.3 Internal Consistency Analysis of the Adopted Scales

An internal consistency analysis, which allows us to study the properties of the measure scales and the questions they comprise, was carried out in Part B of the questionnaires.

The adopted scale was an ordinal *Likert* type scale, with five answer alternatives (from "1" to "5") ranging from "Completely disagree" to "Completely agree". It included 15 items, organized into 4 factors (Table 5).

In the first dimension, Factor 1 – ER, *Cronbach's Alpha* value is over 0.80, and the unidimensional data can be considered adequate (Table 6).

In Factor 2 – Impact of the school-family relationship in the teacher-learning process, *Cronbach's Alpha* is over 0.70, and the unidimensional data can be considered acceptable (Table 6). In Factor 3 – TR, *Cronbach's Alpha* value comes close to 0.70, and the unidimensional data can be considered acceptable (Table 6). Finally, in Factor 4 – School-family relationship difficulties, *Cronbach's Alpha* is below 0.70, and the unidimensional data cannot be considered acceptable (Table 6).

The scale used in Part B – "Student Report", is suitable to measure the factors that were constructed for this sample, with the exception of factor 4; despite this limitation the scale was nevertheless used.

Additionally, we asked the Teachers and Guardians to indicate the advantages and disadvantages of using an ER. Suggested advantages by Teachers and Guardians concerning the use of an ER amounted to 84.7 % and 84.3 %, respectively. As to disadvantages, we registered 24.3 % for Teachers and 20.2 % for Guardians. The most frequently referred advantages in the adoption of an ER, by Teachers and Guardians, were: "Faster, more efficient and safer communication"; "Saves paper"; "Decreases repetitive tasks"; "Customized access to information, at any given time and place";

Table 5. Organization of factors and respective items.

Factors	Items
Factor 1 – ER	9. It is important to find a tool that will allow for the intensification of and an easier communication between the School and the Guardian
	10. The adoption of an Electronic Report may contribute to an improved communication between the school and the Guardian
	11. An Electronic Record may constitute an important work tool for Class Directors, Teachers, Parents and School Principals
	12. An Electronic Report may constitute a communication resource between every Teacher and Guardians
	13. An Electronic Report may be used by School Principals to contact Teachers
	14. An Electronic Report may be used by School Principals to contact Guardians
	15. An Electronic Record offers more advantages than the Traditional Record where the communication between the multiple school community actors is concerned
Factor 2 – Impact of the school-family relationship in the teaching-learning process	1. The relationship between the family and the school is fundamental for the teaching-learning process
	3. Whenever the parents are involved the development of the students is potentially improved
	4. The participation of parents in the school benefits the Teachers and contributes to a more constructive work on their part
	5. Parents assume more favourable attitudes towards Teachers when they cooperate constructively with each other
Factor 3 – TR	7. The Traditional Student Record constitutes an important tool in family-school communications
	8. Teachers regularly use the Traditional Record in their communications with Guardians
Factor 4 – School-family relationship difficulties	2. The relationship between the school and the family is not always easy and stands out as a complex relationship
	** 6. Parents can easily head to the school during the receiving hours appointed by Class Directors and in meeting days

Table 6. Cronbach's Alpha results for each factor

Factor 1		**Factor 2**	
Cronbach's Alpha	Nr. of Items	Cronbach's Alpha	Nr. of Items
0.922	7	**0.730**	4
Factor 3		**Factor 3**	
Cronbach's Alpha	Nr. of Items	Cronbach's Alpha	Nr. of Items
0.661	4	**0.661**	4

"Daily follow-up" and "Virtual Presence". In their answers given to the question involving the disadvantages of adopting an ER the respondents were objective, indicating the inexistence of disadvantages, the possible lack of internet access by some and, mainly, the lack of computer skills.

Part C of the Guardians and Teachers questionnaire focused on the inclusion of items in an ER. We observed that 95 % of the Guardians and 94 % of the teachers believe that all the items of the TR should be included in an ER.

We asked the Guardians and the teachers to indicate their level of agreement as to the inclusion of the following items in a prospective ER, as shown in Table 7.

From the obtained answers, all items reveal a high average agreement, higher for statements 3, 1 and 2 followed by 4, 6, 5 and 7.

From the answers given by the Teachers in our study sample, the items that reveal a higher average agreement involve the statements 2, 3, 4, 6 and 1 followed by 5, and lower for 7. We also asked our study participants to identify some of the items that, in their view, should be included in an ER. Faced with multiple responses, we opted for the inclusion of the most significant, namely the record of absences, mid-term evaluations, record of school entries and exits and communication between Guardians.most welcome.

3 Presentation and Discussion of the Results Concerning the Investigation Questions

The developed study was based on a set of questions included both in the questionnaire given to Guardians and in the questionnaire given to Teachers, which were subject to analysis.

The first investigation questions, "School-family communication is fundamental for the teaching-learning process", relates to Factor 2 – Impact of the school-family relationship in the teaching-learning process, and comprises the following statements: 1, 3, 4 and 5 (Table 5).

The majority of Guardians (96.7 %) believes that the relationship between the school and the family is fundamental for the teaching-learning process. This is also the perception expressed by the majority of Teachers (99.3 %).

Table 7. Inclusion of items in a prospective ER

Items
1. Absence justifications by the Guardians
2. Absence management by the Class Director
3. Mid-term evaluation record
4. Evaluation criteria and tools for each subject
4. Student self-assessment forms
5. Communication management (Messages, SMS and Emails)
6. Videoconference

The observed average values reveal the presented variations, with all items showing a high level of agreement, both for Guardians and Teachers. Factor 2 – Impact of the school-family relationship in the teaching-learning process, thus reveals an average value of 4.49 for Guardians and of 4.55 for Teachers, which come significantly close to the maximum possible value. In conclusion, we can say that both for Guardians and Teachers, the communication between the school and the family is fundamental for the teaching-learning process.

The analysis of the second investigation question "The TR is a fundamental tool in school-family communication" relates to Factor 3 – TR, and comprises statements: 7 and 8 (Table 5). Teachers regularly use the student TR when communicating with Guardians. Indeed, both Guardians (88,8 %) and Teachers (87,9 %) perceive the student TR as an important school-family communication tool.

The observed average values reveal the presented variations, with all items showing a high level of agreement, both for Guardians and Teachers. Factor 3 – TR, thus reveals an average value of 4.20 for Guardians and of 4.13 for Teachers, which come significantly close to the maximum possible value. In conclusion, we can say that both for Guardians and Teachers, the TR constitutes a fundamental school-family communication tool.

The analysis of the investigation question "The ER improves school-family communication", relates to the analysis of Factor 1 – ER, and comprises statements: 9, 10, 11, 12, 13, 14 and 15 (Table 5).

In our sample, Guardians (83.9 %) and Teachers (92.3 %) are unanimous as to the relevancy of finding a tool that allows for the intensification of and an easier communication between the school and the Guardian. Guardians (78 %) and Teachers (87.3 %) go so far as to admit that an ER may become a communication and work resource, between the school and the family. The majority of Guardians (73 %) and Teachers (82 %) believe that the ER may bring more benefits when compared with the TR.

The observed average values reveal the presented variations, with all items showing a high level of agreement, both of Guardians and Teachers. Factor 1 – ER, thus reveals an average value of 4.11 for Guardians and of 4.33 for Teachers, which come significantly close to the maximum possible value.

In conclusion, we can say that both Guardians and Teachers believe that the ER promotes and improves the communication between the school and the family.

The herein presented measures of dispersion reveal relatively low values for the standard deviation, with values below one or coefficient of variation percentages of twenty or less (Tables 8, 9 and 10). Therefore, everything suggests a good dispersion of the obtained results, pointing to a greater homogenization of answers and, consequently, an average value that is more representative of reality.

Table 8. Measures of Dispersion of Factor 2 – Influence of the School/Family relationship in the teaching-learning process and its items.

	Guardians				Teachers			
	N	Average	Standard deviation	Coefficient of variation	N	Average	Standard deviation	Coefficient of variation
Factor 2 – Influence of the school-family relationship in the teaching-learning process	**1002**	**4,49**	**0,54**	**12 %**	**300**	**4,55**	**0,47**	**10 %**
1. The relationship between the school and the family is fundamental for the teaching-learning process.	1002	4,79	0,53	11 %	300	4,82	0,44	9 %
3. Whenever the parents are involved, the development conditions of the students are potentially improved	1001	4,37	0,79	18 %	299	4,47	0,74	17 %
4. The participation of the parents in the school benefits the Teachers, and contributes to a more constructive work on their part	1001	4,34	0,79	18 %	299	4,31	0,77	18 %
5. Parents assume more favourable attitudes towards Teachers when they cooperate constructively with each other	1001	4,47	0,74	16 %	299	4,60	0,61	13 %

Table 9. Measures of Dispersion of Factor 3 – TR.

	Guardian				Teacher			
	N	Average	Standard deviation	Coefficient of variation	N	Average	Standard deviation	Coefficient of variation
Factor 3 – Traditional Record	**1002**	**4,20**	**0,83**	**20 %**	**300**	**4,13**	**0,73**	**18 %**
7. The Traditional Student Record constitutes an important tool in family-school communications	1002	4,49	0,82	18 %	299	4,33	0,79	18 %
8. Teachers regularly use the Traditional Record in their communications with Guardians	1002	3,91	1,08	28 %	300	3,94	0,91	23 %

Table 10. Measures of Dispersion of Factor 1 – ER.

	Guardian				Teacher			
	N	Average	Standard deviation	Coefficient of variation	N	Average	Standard deviation	Coefficient of variation
Factor 1 – Electronic Report	**1002**	**4,11**	**0,85**	**21 %**	**300**	**4,33**	**0,71**	**16 %**
9. It is important to find a tool that will allow for the intensification of and an easier communication between the school and the Guardian	1002	4,32	0,91	21 %	299	4,57	0,73	16 %
10. The adoption of an Electronic Report may contribute to an improved communication between the school and the Guardian	1002	4,03	1,09	27 %	300	4,35	0,81	19 %
11. An Electronic Record may constitute an important work tool for Class Directors, Teachers, Parents and School Principals	1002	4,12	1,02	25 %	298	4,43	0,81	18 %

(*Continued*)

Table 10. (*Continued*)

	Guardian				Teacher			
	N	Average	Standard deviation	Coefficient of variation	N	Average	Standard deviation	Coefficient of variation
12. An Electronic Report may constitute a communication resource between every Teacher and Guardians	1001	4,13	1,01	24 %	300	4,37	0,82	19 %
13. An Electronic Report may be used by School Principals to contact Teachers	1002	4,05	1,01	25 %	300	4,04	1,14	28 %
14. An Electronic Report may be used by School Principals to contact Guardians	1002	4,13	1,00	24 %	300	4,28	0,90	21 %

4 Conclusion

The collaboration between the school and the family, two fundamental institutions for the education of the child, is essential to improve the teaching-learning process, and promotes the desired educational success. Education is a task that must be shared by Teachers, Guardians and community institutions [19]. Because we believe that, in the teaching-learning process as a whole, communication and, particularly, the means with which we communicate, are essential to educational success, we directed our study to understand the opinion of Guardians and Teachers as to the feasibility of adopting an ER in the school-family relationship.

As a starting point for the development of this study we posed three hypothesis: H1: The educational community values communication between the school and the family; H2: The TR is a fundamental tool in school-family communication; H3: The ER improves school-family communication.

We concluded, based on the collected answers, that Teachers and Guardians perceive school-family communication as being fundamental in the teaching-learning process, to the extent that the factor concerning the impact of the school-family relationship in the teaching-learning process obtained an average of 4.49 amongst Guardians and 4.55 amongst Teachers, which comes significantly close to the maximum value possible. For Teachers and Guardians, the TR constitutes a fundamental tool in school-family communications. Currently, the TR is, for Guardians, the only vehicle of communication with Teachers, as they cannot head to school during opening hours. From the point of view of the Guardians and Teachers who participated in the study, the ER can promote and improve school-family communications. Indeed, the analysis carried out to the answers collected from our study participants led us to conclude that the Factor 1 – ER, reveals an average value of 4.11 amongst Guardians and of 4.33

amongst Teachers, which come significantly close to the maximum value possible. The majority of Teachers and Guardians asserts that the ER should include the items that are already available in the TR. They also stress the importance of including the following items: record of absences, mid-term evaluations, record of entries and exits from school, as well as communication amongst Guardians.

Considering the obtained results, we observed that the three hypothesis were reinforced, as the majority of the respondents (Guardians and Teachers), believe that the communication between the school and the family is fundamental for the teaching-learning process; moreover, the TR emerges as fundamental tool in school-family communications and the ER promotes and improves this school-family communication.

Therefore, the ER may assume, in a ground breaking, dynamic and assertive way, a facilitating role in school-family communications, allowing for the development of partnerships.

In this sense, the development of a web application, the ER, is justified, to enable the consultation of the entire body of information concerning the student and promote an easier, quicker and safer communication between the school and the family.

References

1. Abrantes, B.: Conceção e Desenvolvimento de um Ambiente de Aprendizagem Pessoal Baseado em Ferramentas Web 2.0. Universidade de Aveiro (2009)
2. Bartlett, M.S.: The effect of standardization on a chi square approximation in fator analysis. Biometrika **38**, 337–344 (1951)
3. Berto, R.M.V.S., Nakano, D.N.: A Produção Científica nos Anais do Encontro Nacional de Engenharia de Produção: Um Levantamento de Métodos e Tipos de Pesquisa. Revista Produção **9**(2), 65–76 (2000)
4. Boonen, A.: Pourquoi utiliser les technologies de l'information et de la communicationdans le domaine de l'éducation? In: Scheffknecht, J.J. (ed.) Les technologies de l'informationà l'école: raisons et stratégies pour un investissement. Strasbourg, Conseil de l'Europe (2000)
5. Costello, A., Osborne, J.: Best practices in exploratory fator analysis: four recommendations for getting the most from your analysis. Pratical Assess. Res. Eval. **10**(1), 1–9 (2005)
6. Groves, R.M., Brick, M.J., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., Tourangeau, R.: Survey Methodology. Wiley, Hoboken (2004)
7. Guimarães, R.C., Sarsfield Cabral, J.A.: Estatística, 2ª Edição. Verlag Dashöfer (2010)
8. Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C.: Análise multivariada de dados, 5th edn. Bookman, Porto Alegre (2005)
9. Jiménez, E.G., Flores, J., et al.: Cuadernos de estadística: análisis factorial. La Muralla, Salamanca (2000)
10. Kaiser, H.F.: An index of fatorial simplicity. Psychometrika **39**(1), 31–36 (1974)
11. Leong, F.T.L., Austin, J.T.: The Psychology Research Handbook, 2nd edn, p. 516. Sage Publications, Thousand Oaks (2006)
12. Lima, L., Sá, V.: A participação dos pais na governação democrática das escolas. In: Lima, J. (Org.). Pais e professores, um desafio à cooperação. Edições Asa, Porto (2002)
13. Lourenço, L.P.R.: Envolvimento dos Encarregados de Educação na Escola: Conceções e Práticas. Universidade de Lisboa (2008)

14. Marques, R.: Colaboração Escola-Famílias: um conceito para melhorar a Educação. Ler Educação, 8 (1992)
15. Marques, R.: A escola e os pais: como colaborar?. Texto Editora, Lisboa (1993)
16. Mattelart, A.: The Information Society. Sage Publications, London (2003)
17. McLean, N.: Technology can bridge the gap between parents and schools. The Independent, 5 November (2009)
18. Montadon, C., Perrenoud, P.: Entre pais e professores, um diálogo impossível? Para uma análise sociológica das interações entre a família e a escola. Celta, Oeiras (2001)
19. Paro, V.: Qualidade do Ensino: A Contribuição dos Pais. Xamã, São Paulo (2003)
20. Pereira, A.: Análise de dados para Ciências Sociais e Psicologia. Edições Sílabo, Lisboa (2006)
21. Pereira, C.B.: O marketing do lugarzinho: uma aplicação exploratória da técnica de índice de preços hedônicos a jovens consumidores de restaurantes na cidade de São Paulo. Tese de Doutoramento em Administração – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo. São Paulo, 165 (2004)
22. Salvador, P., Rocha, Á.: An assessment of content quality in websites of basic and secondary portuguese schools. In: Rocha, Á., et al. (eds.) New Perspectives in Information Systems and Technologies. Advances in Intelligente Systems and Computing 275, vol. I. Springer International Publishing, Switzerland (2014)
23. Sarmento, T.: (RE)pensar a interação escola família. Revista Portuguesa de Educação **18**, 53–75 (2005)
24. Sarmento, T., Marques, J.: A Escola e os Pais, Coleção Infans. Centro de Estudos da Criança, Braga (2002)
25. Silva, P.: Escola-família: Tensões e potencialidades de uma relação. In: J.Á. de Lima (Org.) Pais e Professores, Um Desafio à Cooperação. Porto, Edições ASA (2002)
26. Stevens, J.: Applied Multivariate Statistics for the Social Sciences, 3rd edn. Lawrence Erlbaum, Mahway (1996)
27. Tabachnik, B.G., Fidell, L.S.: Using Multivariate Statistics, 5th edn. Pearson Education, Boston (2006)
28. Villas-Boas, M.: A parceria entre a escola, a família e a comunidade: reuniões de pais. Edição do Departamento de Avaliação Prospetiva e Planeamento do Ministério da Educação, Lisboa (2000)
29. Villas-Boas, M.: Escola e família: uma relação produtiva de aprendizagem em sociedades multiculturais. Escola Superior João de Deus, Lisboa (2001)

Cryptocurrencies as a Disruption? Empirical Findings on User Adoption and Future Potential of Bitcoin and Co

Aaron W. Baur(✉), Julian Bühler, Markus Bick, and Charlotte S. Bonorden

ESCP Europe Business School Berlin, Berlin, Germany
{abaur, jbuehler, mbick}@escpeurope.eu,
charlotte.bonorden@edu.escpeurope.eu

Abstract. In this paper, we examine cryptocurrencies as a potentially disruptive sort of payment method. Due to its relative importance, we focus in particular on Bitcoin. Through an inductive, exploratory interview approach with 13 individuals in three distinct groups, the determinants usability, usefulness, and subjective norm that could make Bitcoin a game-changer are explored. The results reveal that most stakeholders consider perceived ease of use still rather low, with perceived usefulness varying according to the user group. The notion of Bitcoin as having much future potential as a payment method is confirmed across all interviewees. Interestingly, the underlying concept of a blockchain is also seen as a potential revolutionary way to create a more just society based on open platforms and open data. However, the reasons of why Bitcoin is actually a disruption to existing solutions varies widely.

Keywords: Electronic payment · Bitcoin · Cryptocurrency · Digital currency · e-commerce · Technology adoption · TAM · Blockchain

1 Introduction

The Internet and the massive growth of e-commerce have bred various new online payment and money transaction methods in the last several years. This is being spurred by a new hype in 'fin-tech' startups in the global technology hubs. Even though famous quotes like "banking is necessary, banks are not" made by former Wells Fargo CEO Richard Kovacevich [1] are already quite dated, technology finally seems ready to enable real innovations in this sector. A new generation of founders is motivated to revolutionize how the financial industry works, what society thinks about the concept of money, and how the future of monetary transactions will look.

As the number of alternative payment types has grown, traditional means of online payment, like credit cards, are on the downgrade as they are fighting with competition that offers higher user-friendliness, more security, and lower costs [2, 3]. As Koley [4] reports, even though the virtual duopoly of Visa and MasterCard is still experiencing healthy growth rates (9.5 % and 9.6 %, respectively) and remains the preferred method for (offline) point-of-sale (POS) transactions, firms like JCB and UnionPay outperform

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 63–80, 2015.
DOI: 10.1007/978-3-319-25013-7_6

Rank	Name	Market Cap
1	Bitcoin	$ 3,372,991,146
2	Ripple	$ 213,123,917
3	Litecoin	$ 56,741,480
4	Dash	$ 15,396,138
5	Stellar	$ 13,645,326
6	Dogecoin	$ 11,934,891
7	BitShares	$ 9,554,346
8	Nxt	$ 9,217,893
9	BanxShares	$ 7,508,688
10	Peercoin	$ 5,772,497

Fig. 1. Top ten cryptocurrencies and market cap in USD, as of May 15, 2015 (Source: http://coinmarketcap.com/currencies/views/all/)

them by far (growing 20.7 % and 44.8 %, respectively). This provides hints of altered consumer behavior including a perceived viable alternative to credit cards and a gradual change in the payment market [5]. PayPal, which was originally meant for facilitating eBay transactions, is now one of the major players in online payment [6]. Additionally, new firms that are home to completely different industries than banking, like Google and Apple, offer solutions that have already gained a remarkable level of user acceptance and dissemination among society and businesses alike [7, 8]. However, all of these solutions remain tied in one way or another to the user's regular bank account and a traditional notion of the concept money.

Another, potentially far more disruptive innovation in this field applies to crypto-currencies [9]. Advocates of these new means of payment claim various advantages, like a fully decentralized, peer-to-peer transaction system, elimination of chargeback risks, lower associated transaction costs, increased level of security, greater ease of use, and full support for mobile devices [10, 11]. Of the several hundred different crypto-currencies,[1] Bitcoin is the most well-known, most discussed, and most widely traded one with a market capitalization of more than three bn USD (see Fig. 1 for the top ten cryptocurrencies as of May 15, 2015). Hence, in this paper we focus on Bitcoin and use 'Bitcoin' and 'cryptocurrency' somewhat interchangeably.

The body of scientific research in the field of cryptocurrencies is still very manageable (see Related Work section below). One stream that especially lacks work is the question of user adoption and the relevant drivers of it (e.g., [12, 13]). Academic research has mostly neglected the user perspective, which connects the technological infrastructure with the established economies [12]. In particular, the decisive impact on society as a whole that users expect of these blockchain technologies has not been gathered. To the best of our knowledge, there has not been any empirical research applying detailed interviews to gain 'rich knowledge' about these kinds of questions. With this paper, we contribute to this under-researched turf by means of an exploratory, qualitative approach using 13 semi-structured interviews.

[1] The website https://coinmarketcap.com lists 560 as of May 15, 2015.

In particular, the study aims to examine the following research questions:

1. What are the perceived advantages and disadvantages of Bitcoin as compared to other forms of (electronic) payment?
2. What are the drivers and barriers Bitcoin users, merchants, and experts see in the adoption of Bitcoin?
3. How do users, merchants, and experts evaluate the future potential of Bitcoin as a serious means of currency, unit of account, asset, and disruption for society?

The remainder of this paper is structured as follows: After giving an overview of the applicable literature on cryptocurrencies and present payment options, the applied research methodology is presented. In the following part, the interview findings are discussed. The paper closes with a conclusion section and also discusses contributions, limitations, and future avenues to advance this research area.

The aim is to explore Bitcoin's main advantages and disadvantages compared to current payment types, to identify the drivers and barriers of Bitcoin's adoption, and to evaluate its future potential as seen by the three distinct groups of (pure) users, merchants, and experts.

2 Related Work

In order to find empirical works on Bitcoin adoption and usage, we conducted a thorough literature search following the frameworks of vom Brocke et al., Levy and Ellis, and Webster and Watson [14–16]. Search terms were 'Bitcoin', 'crypto AND currency[ies]', and 'cryptocurrency[-ies]'. As suggested by Chen et al. [17], we searched the databases *Web of Science, EBSCO Business Source Complete, IEEE Xplore, ScienceDirect*, and the *ACM Digital Library*. To be as exhaustive as possible, the important IS conferences ICIS, ECIS, HICSS, PACIS, and AMCIS were also explored with the identical keywords. As the 'mother paper' of Bitcoin by Satoshi Nakamoto (most likely a group pseudonym) was just released in 2008 [18], we did not limit the search timeframe. In fact, it was not until 2011 that research papers about cryptocurrencies were published in journals and from conferences. From 2013, larger peer-reviewed journals have been picking up the topic and have started to accept papers more frequently.

The results, combined with the work of Scott [19] and Böhme et al. [20], revealed four main streams of research in regard to Bitcoin, addressing three different layers:

- *Technical:* This stream includes technical details about the 'back-end' of cryptocurrency, like cryptography, mathematical models, and system design. It consists of two sub-streams: On the one hand, conceptual and prototyping work in regard to the *protocol* layer (e.g., [21–23]), and on the other, research concerning the *network* layer, i.e., mainly security-related works made up of experimental and quantitative research designs (e.g., [24–26]).

The other three streams address the *ecosystem* layer:

- *Economic:* This body of research looks at cryptocurrency from an economist's perspective, and mainly includes traditional economic models, portfolio theory, incentive structures, and the like (e.g., [20, 27–31]).
- *Regulatory:* Here, researchers discuss legal, fiscal, tax, and regulatory issues of Bitcoin as a new form of currency (e.g., [10, 32–34]).
- *Social Science:* Along this road, this research is about sociology, trust, anthropology, ethics, and politics in regard to the new phenomenon of cryptocurrencies (e.g., [35–38]).

Surprisingly, the *Social Science* branch is the least developed, but probably the most relevant for Information Systems (IS) research. Hence, this paper focuses on this human-centered aspect of cryptocurrency.

2.1 Traditional Digital Payment Solutions

For a long time, credit cards have been the dominant means of payment on the Internet. Credit cards like Visa, MasterCard, American Express, and Discover still denote the highest volumes of all forms of payments. For instance, daily transaction volumes of Visa amount to almost 18 bn USD (Fig. 2). Due to certain drawbacks of credit cards to merchants and consumers—e.g., the possibility of chargebacks, high fees, and a significant fraud risk—a wide array of alternative means of online payment has developed. Important players include China Union Pay, PayPal, and increasingly Bitcoin, albeit with a huge gap when compared to their old-school credit card competitors.

New players from the 'fin-tech' startup realm enter the market and push the financial market to become more digital and more efficient, encouraging higher transaction volumes and enhancing security. Payment service providers advertise better security standards, a better ease of use, and lower fees. However, none of the current alternatives could so far solve the three inherent major flaws completely.

- Payments through alternative providers still run through customers' credit/debit cards or bank accounts, so transaction fees cannot be cut entirely, and the savings are sometimes even marginal. Furthermore, cross-border transactions still involve significant fees, regardless of whether the money is transferred directly via credit card or via a service like PayPal or Skrill.

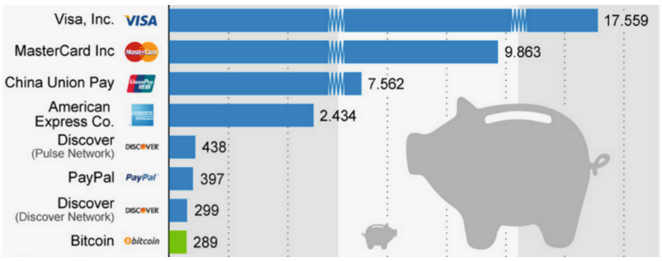


Fig. 2. Daily transaction volume in mn USD, as of the end of 2013 (Source: http://statista.com)

- The risk of fraud is still present, as every time a customer purchases goods online, his or her personal data like name, address, and financial information are transferred over the Internet. This information therefore remains prone to identity theft, no matter whether it is stored only at one single point (PayPal) or at several.
- Usability or ease of use is still a major weakness, especially concerning older people who would like to purchase goods online.

These flaws and other ideas have led to a completely new approach to e-payment: cryptocurrencies.

2.2 Overview of Cryptocurrency

From a technical perspective, cryptocurrency can be defined as a "type of digital currency which relies on cryptography, usually alongside a proof-of-work scheme, in order to create and manage the currency. A decentralized network of peer-to-peer computer nodes working in sync creates and verifies transactions of transfer of said currency within the network" [39]. A more practical definition is given by Ametrano [40], stating that cryptocurrency "can be transferred instantly and securely between any two parties, using the Internet infrastructure and cryptographic security with no need for a trusted third party. Its value is not backed by any single government or organization." Furthermore, Wiatr [41] gives the definition that "a cryptocurrency is a modern digital medium of exchange. It is a new decentralized, limited and peer-to-peer payment system. Most cryptocurrencies are created to introduce new units of currency, whose total amount is limited. All cryptocurrencies use cryptography to control the creation and transfer of money." Summarizing the above, we basically follow Spenkelink [42] and find four main characteristics of cryptocurrencies. First, cryptocurrency works through a decentralized network, meaning that it is free of any external regulations. Second, it has a strong P2P-approach. This guarantees the direct transaction of money between two parties, regardless of whether it is B2B, B2C, or C2C. Third, it uses the public Internet, which provides speed, efficiency, and resilience. Finally, cryptocurrency uses public-key cryptography to make money transactions as secure as possible.

After the introduction of Bitcoin as the first digital currency in 2009, the number of cryptocurrencies began to rise, and it quickly became a global phenomenon [43, 44]. Besides Bitcoin, there are hundreds of alternatives, each with different strengths. For example, *Litecoin* uses the *scrypt* encryption and claims to transact four times faster than Bitcoin, making it a viable alternative for high-speed applications like financial trading [20]. The so-called *Altcoins* can be regarded as 'forks' of either Bitcoin or Litecoin, since they are based on the same type of algorithms. Examples include *Feathercoin*, *ChinaCoin*, and *Dogecoin* [39].

Even though in theory Bitcoins can be mined, i.e., created by everyone, the system behind Bitcoin ensures certain characteristics that are important conditions for valid (fiat) currencies. First, scarcity: Bitcoins are limited to 21 million, keeping inflation low [45]. Second, security: Bitcoin follows the highest security standards using public and private keys [46]. After a person registers as a Bitcoin provider, he or she gets a digital

wallet identification number (ID) that will be shown in public, once the person transfers Bitcoin. All past transactions are recorded in the blockchain. Third, simplicity: Transferring Bitcoins is easy, requiring a Bitcoin application or wallet on a mobile device and the required amount of Bitcoins, normally received through a Bitcoin exchange or an ATM.[2] The transfer then only takes minutes, regardless of geographical distance.

2.3 Value-Add and Risks of Cryptocurrencies

Most authors support the argument that Bitcoin has a significant future and will become a larger threat for credit card companies [2, 47, 48], but for now the potential is estimated to be limited to only specific fields [13, 49]. Grinberg [11] underlines the fact that Bitcoin is likely to be the new, revolutionary payment method for micropayments, because those are hardly profitable when credit card fees are applied to such small notional amounts. Böhme et al. [20] argue that it is becoming less certain whether the design of Bitcoin meets the requirements to replace credit card payments for the everyday consumer, particularly with regard to security standards. Furthermore, they illustrate the fact that there are numerous competing virtual currencies, for instance Litecoin that accelerates payment confirmations, or 'NXT' that reduces computational requirements. To become an essential part of the customer's daily life, Bitcoin first needs to achieve the overall reliance in value offer and user adoption.

Spenkelink [42] follows a broader perspective by concluding that the future of cryptocurrencies is very unclear in general and one can only speculate about its development. He further claims that many different usage scenarios and different stakeholder needs prevail, which makes the estimation of Bitcoin's mass adoption even harder. Descôteaux [50] holds the opinion that for the future development of Bitcoin, an appropriate legal and regulatory framework is extremely important in order to overcome the existing risks and threats slowing down the adoption of cryptocurrency. The literature also discusses the drawbacks and disadvantages of Bitcoin, such as the possibility of money laundering [51–54], trade in various illegal goods [55, 56], potential financing of terrorism [51, 57], Bitcoin loss or theft [21, 58], and tax evasion or enormous volatility [27, 43, 59]. These aspects were also mentioned by our interview partners.

3 Research Methodology

As described above, only limited research on Bitcoin's advantages, disadvantages, disruptive potentials, and adoption drivers exists. Therefore, we approached the research by following an entirely open, inductive, exploratory research design [60, 61]. Accordingly, an interview guideline was designed based on the literature review, but also including very open questions to motivate interviewees to come up with their own ideas, thoughts, anecdotes, and reasoning. To get insights into how different groups see

[2] The website http://coinatmradar.com/ lists 389 ATMs globally as of May 15, 2015.

Table 1. Sample description

Individual	Group	Occupation	Bitcoin Experience	Means of Interview
C1	Consumer	Student	none	Face-to-Face
C2		Entrepreneur	medium	Face-to-Face
C3		Entrepreneur	extensive	Skype
C4		Scientist	none	Face-to-Face
C5		Architect	none	Face-to-Face
M1	Merchant (E-Commerce)	Innovation Manager, Retail Industry	extensive	Skype
M2		CEO, Wholesale Industry	none	Face-to-Face
M3		Head of Digital, Media Industry	little	Face-to-Face
M4		Market Researcher, Retail Industry	none	Skype
B1	Bitcoin Exchange	CEO	extensive	Skype
B2		CTO	extensive	Skype
B3		Senior Marketer	medium	Face-to-Face
B4		Accountant	medium	Face-to-Face

and experience Bitcoin, the guideline was used to conduct semi-structured interviews in November and December 2014. To gain maximum diversity and coverage, 13 interviews with individuals from three distinct groups were used as a sample [62, 63]. These are end-consumers, e-commerce merchants, and employees of Bitcoin exchanges with differing levels of experience with Bitcoin (see Table 1 for the sample description). The selection of the interview partners was based on a purposive sampling strategy [64]. This ensured the inclusion of a wide array of aspects of the evidence gathered from informants [65, 66]. The interviews were recorded and subsequently transcribed [66]. They lasted between 34 and 56 min.

Interview transcripts were analyzed using the open-coding technique [67]. Relevant passages were identified and axial coding was subsequently used to refine the categories. The analysis was diligently carried out to fulfill the common requirements for qualitative research (e.g., [68–70]). The interpretation process was rule-driven and documented ("procedural validity"; [70]) In addition, we triangulated the data [70] by interviewing people with different perspectives on and knowledge about the phenomenon (i.e., three distinct user groups). Therefore, the results can be assumed to be reliable [71].

4 Analysis and Results

While analyzing the interview transcripts in order to find answers to the research questions, we decided to apply the established constructs *perceived ease of use* and *perceived usefulness* of the Technology Acceptance Model TAM [72, 73] and *subjective norm* of its successor, TAM2 [74, 75], as they work well as ordering categories.

Another one, *future potential*, emerged out of the data. Table 2 summarizes the findings, which are organized around the four research constructs and the case codes. We have included original quotes, which is the 'beauty of qualitative research' [76] and more often than not provide the deepest insight into the matter. In the following paragraphs, we discuss the condensed quintessence of our findings.

4.1 Perceived Ease of Use

Consumers found mobile wallets and Bitcoin in general difficult to use, therefore requiring some training, although one interviewee regarded Bitcoin as easy and self-explanatory. As buying, selling, or sending Bitcoin is very similar to using a Google Wallet, it is possible that more consumers will come across technical difficulties when using Bitcoin for the first time. They agreed that especially the younger generation

Table 2. Findings and case coding

Research Constructs	Codes	Opinions of Individuals C = Consumer M = Merchant B = Bitcoin Exchange	Exemplifying Quotes
Perceived Ease of Use	Mobile Wallets	useful, but cumbersome (C1/2/3); needs more development and refinement (C2); Bitcoin best use case for wallets (B3)	**C2:** "Google Wallet was really complicated and not easy to use" **C5:** "I have not yet dared to install some kind of Wallet"
	Usage / Implementation Process	easy and self-explanatory (C2); potentially easy (C1/C4/M2); easy and fast, hard at POS (M1/3); easy, but not easy enough for mainstream (B1/2/3)	**C1:** "sounds pretty easy" **B1:** "the general mainstream user has no clue about Bitcoin"
	Technical Understanding (Explanation)	easy handling (C2); merchants also need training (B2); easy, but concept and idea not yet understood by society (B1/4)	**M1:** "not a technical, but a real societal challenge"
	Capability (Education/Age)	some learning necessary (C3); only accepted by generation Y/Z (B3/4)	**C3:** "you have to learn how to use it" **C1:** "certainly, the under 40 crowd is far more likely to embrace it, if it is convenient in the end"
	Perceived Risks	volatility (M1/2); advances in technology and usability of other mobile payments (B1); security and storage (B2); use as speculative trading device (B1/2/3)	**B3:** "the volatility of Bitcoin and the frictions of moving money in and out of Bitcoin are a limiting factor in its adoption" **B2:** "Bitcoin also comes with the burden of keeping it secure and accessible at the same time. No really good solution exists yet that solves both problems satisfactorily"

(*Continued*)

Table 2. (*Continued)*

	Convenience	cuts out annoyance of searching for cash at POS (C1); Bitcoin more convenient than rest (C2/3); with QR code very convenient (B1/3)	**M1:** "convenient mobile payments arose at the expense of security standards" **B1:** "paying with Bitcoin is both convenient and secure. We just need to scan a QR code in these days and our payments are transferred successfully. Anywhere, anytime" **B2:** "just scan the QR code, done!"
Perceived Usefulness	Anonymity	big problem with CCs, Bitcoin may be solution (C1/2/3); anonymity not an issue for consumers, only important for special target industries (M1/2/3, B1/2/3/4)	**C2:** "people are more and more paranoid about their privacy" **M1:** "maybe relevant in gambling and sex business" **M2:** "higher anonymity will not increase payment share of Bitcoin"
	Security / Fraud	involved in CC fraud (C1), high perceived level of security with Bitcoin (C1); 100% security not possible (all); fraud of CCs not main argument (M1/2/3); security a very strong Bitcoin bonus (B1/2/3)	**C1:** "American CCs are not as secure as European ones that have chips in them, we have worse technology than Mongolia" **M1:** "CC fraud covered by insurance" **M2:** "Bitcoin accounts can be hacked as well and used for transactions, hence fraud risk still exists" **B2:** "blockchain technology serves as a trust anchor for both parts of a trade"
	Transaction Fees	alternatives to CC highly needed (C2); lower fees most persuasive argument for Bitcoin (C1), wiring fees exorbitant (C1); payout in local currency will have high fees (M1); low fees essential (M1/2/3); esp. relevant to merchants (B1/2/3)	**C2:** "airlines ripping people off when paying with CC, very frustrating" **M1/2/3:** "most promising argument for Bitcoin" **M3:** "interesting for micropayments" **B1:** "transaction costs have 'skyrocketed' in these days and hence leave merchants from certain industries at a competitive disadvantage"
	Instant Transactions	more relevant to merchants (C1/2/3); instant transaction very helpful (M3/4); time and no chargebacks strong arguments pro Bitcoin (B1/2/3)	**M2:** "receiving payout faster from customers to us has its benefits and we would prefer it" **M3:** "time is money, so great" **B1:** "for merchants, Bitcoin displays a direct money inflow of their revenues, whereas with CC payment they need to wait for around a week to receive their payments"
	Global Reach & Cross-Border	great not to search for ATM abroad (C1/3); Bitcoin attractive for remittance (C3); international speed and simplicity very persuasive (B1/2/3)	**B1:** "SEPA has failed, a technology based system like Bitcoin needed to increase speed and decrease fees at the same time" **B2:** "transfer money everywhere in the world at velocity of sound. Payment providers like Western Union have made that possible too, however at high costs"

(*Continued*)

Table 2. (*Continued*)

	Trading / Investment	Bitcoin as a trading currency (C1/2); if not mainstream payment, then alternative investment tool (B2/4)	**B2:** “a secure and efficient, trustful decentralized monetary transfer system with an inherent currency unit that is of non-governmental and non-corporate origin, perfect for investment”
	Emerging Economies	high attractiveness for developing world (all); challenges very different to our market (C2/3); no political restrictions, embargos, capital controls (B2/3)	**B2:** “highly necessary for countries with a high percentage of underbanked people. Why shouldn’t a web shop be able to sell products wherever it can ship it? Most major package delivery services have a wider reach than PayPal or MasterCard”
Subjective Norm	Innovation / Technology	Bitcoin on the forefront (C2); makes mobile wallets attractive (C1/3); Fin-Tech innovations very important (C1/2/3, B1/2/3); innovation reason nr 1 for Bitcoin (M1); innovation crucial for us (M1/2/3)	**M1:** "innovation is all we have, Bitcoin helps!" **M3:** “in my industry (press and media), innovation is key. We can no longer survive if we don’t go digital. We will definitely continue focusing on becoming more innovative” **B2:** “the financial industry has not come up with a good candidate for an online currency / payment system, so Bitcoin is it”
	Competitive Advantage	competitive pressure not that large yet (M2/3); merchants will feel the pressure soon with higher penetration (B1/2/3)	**M3:** "accepting Bitcoin is a competitive advantage right now, for how long remains to be seen" **M2:** “we are looking into Bitcoin but will not expect to adopt it for another 1-2 years. So far most competitors don't accept Bitcoin as it is still a product for early adopters / innovators and misses scale and acceptance in the market"
	Peer Influence	friend trusted more than journalist (C1/5); p2p character of Bitcoin fosters peer influence (B2)	**C1:** “if someone says that Bitcoin is a very practical, efficient and secure way of doing things, I would check it out” **B1:** "word of mouth is needed in order to spread the word and explaining the beauty behind Bitcoin”
	Latest Trends	trends not of relevance (C1/2); people more conservative with latest trends in financial issues (B2); direct, peer-to- peer transfers may be considered cool (B1)	**M1:** "exciting thing, especially for Techies, but from a bottom line perspective, not relevant yet" **B2:** “the traditional financial system has not been very keen on adopting new technologies and when it comes to money, people tend to be more conservative than with other things”
	Lifestyle / Marketing	Bitcoin great marketing asset (M1/3/4)	**M3:** “if a whole industry does it, it makes sense to adopt it. As for now, we us it, but Bitcoin is not a big thing yet” **B1:** "Bitcoin could become a lifestyle product, people will find it increasingly cool to use"
	Alternative: PayPal	no viable alternative (C1/2); good alternative (all others)	**C1:** "takes time to register" **C2:** "another account to keep track of"
	Alternative:	favorite means of payment	**C1/2:** “typing in the details again and again is

(*Continued*)

Table 2. (*Continued*)

Future Potential	Credit Card (CC)	(C1/2); insecure (B1/2/3); OK (M1/2/3)	annoying" **M1:** "used by majority of costumers"
	Alternative: Others	bank transfer / wire (all); Master-pass / Sofortueberweisung (all but C1/2/3)	**B3:** "other alternatives will of course remain, but become less and less important"
	Current Subjective Market Penetration	not clear who accepts Bitcoins currently (C1/2/3); still a niche form of payment, low demand (M1/3); not really accepted (M2); customer demand drives penetration (B1/3/4)	**C2:** "recognizing more 'Pay with Bitcoin'-buttons on official websites people would boost credibility" **B3:** "current lack of easy-to-use and secure solutions for storing and using Bitcoin without third-party services do not make Bitcoin's advantages apparent to new users"
	Subjective Future Market Potential	payout in local currency will boost acceptance (M1/2); more and more solutions pave way for great potential (B2)	**C2:** "will be the future", "easiest, most secure and fastest way of payment" **M1:** "will become a big thing in the future" **M2/3:** "will be big, but never replace CCs" **B3:** "Bitcoin is a very promising candidate for the currency of the internet"
	Time Until Widespread Adoption	quick adoption (C3); very difficult to guess (all but C3); probably very different speed in different regions of world (M1/2, B1/2/3)	**M1:** "in IT stuff, I stopped giving prognoses" **B1:** "hard to tell, maybe 3 to 5 years, also depending on regulation/taxation/legislation"

should manage to handle Bitcoin and related software/apps, however. These findings are similar to those of Spenkelink [42]. Risks were not mentioned, but instead consumers focused on convenience, especially at offline stores (POS).

One merchant complained about the implementation process at POS and stressed that it's not a technical, but rather a societal challenge to switch to Bitcoin. Implementation in online shops was seen as fast and easy, and similar to other payment options. Most saw volatility as the main threat and agreed that current mobile payment solutions are convenient, but insecure; maybe Bitcoin can be a remedy here.

The representatives of Bitcoin exchanges saw Bitcoin as the best use case for mobile wallets. Even though they judged usage and implementation as easy, they still found it not easy enough. In fact, the whole cryptographic system behind Bitcoin was deemed just too complex to be understood by users and merchants, and it discouraged them from giving Bitcoin a try. However, their own companies and the majority of other exchanges already focus on emphasizing the easy set-up, integration, and usage of Bitcoin. This has been confirmed by a check on the landing pages of Bitpay, Cubits, Coinbase, and Safello.[3] Here, tutorials, videos, and other training and marketing activities help users get started.

[3] https://bitpay.com; https://cubits.com; https://www.coinbase.com; https://safello.com.

4.2 Perceived Usefulness

For consumers, anonymity and security rank very high. They feel threatened by credit card fraud and the old technology used, and consumers hope Bitcoin can be a remedy here. Due to high transaction fees of credit cards and especially money wiring (fees here seen as exorbitant), the lower costs are seen as the major argument for using Bitcoin. The instantaneous character of transactions was seen most relevant to merchants. The fact of doing away with ATM searches abroad once Bitcoin is also widely accepted at POS, and the possibility of doing remittances as well as using Bitcoin as an investment vehicle were all seen as very positive. Finally, Bitcoin was found especially suitable for emerging economies, as they face very different challenges (e.g., lack of availability of traditional banking services).

Merchants had a quite different view on anonymity and fraud: Anonymity was only crucial for special target groups, e.g., in online gambling or the sex business, not for mainstream costumers, thus stating the opposite of Pagliery [57]. The security of Bitcoin was also not necessarily seen as higher than that of credit cards, as Bitcoin accounts can also be hacked, which has happened before. Low transaction fees were the single most important argument for merchants, also in regard to micropayments. Having the payout of orders immediately on their accounts was also seen as very advantageous, as they are well aware that time is money.

Bitcoin exchange staff is more convinced about Bitcoin's usefulness regarding the elimination of fraud. All experts believe that security will become an increasing problem in the near future, which is why more secure payment solutions are necessary, and this is a big benefit of Bitcoin. In terms of anonymity, the necessity of finding ways to get rid of skyrocketing transaction fees, and the beneficial effect of instant transactions, the experts strongly agreed with the merchants. In terms of global and cross-border reach, the managers quoted pure speed and simplicity as persuasive arguments in favor of Bitcoin. Whereas one interviewee judged SEPA as failed, another mentioned Western Union as working, but at very high costs. Bitcoin as investment was regarded as ideal, as there is no government or corporate involvement. This is in strong opposition to the findings of Baek and Elbeck [27] and Garcia et al. [77], who see Bitcoin as far too speculative to be used by a non-professional investor. Lastly, emerging economies were mentioned as a suitable target market, since online shops often times can ship to remote locations, but cannot receive funds from there due to capital market restraints.

4.3 Subjective Norm

In terms of innovation and technology, consumers rated Bitcoin to be very innovative, and it helps making mobile wallets attractive. In terms of peer influence, the interviewees so far unfamiliar with the (Bitcoin) process validated that they would give it a try upon recommendation of a friend. Two consumers denied that they give any relevance to current trends or lifestyle issues, i.e., Bitcoin plays no role here.

Merchants focused on the huge importance of innovation for their business, making this category the most decisive factor to offer Bitcoin payments. However, competitive

pressure itself is currently still manageable, with one merchant confirming that they will look into it, but will probably not introduce it for another 1–2 years. One merchant claims Bitcoin to be a competitive advantage. In terms of trends, one merchant mentioned that it is a cool thing, especially for early-adopters/techies, but has no effect on revenue or profit, i.e., the bottom line. This statement basically confirms the research of Kostakis and Giotitsas [78]. The merchant also claims it to be somewhat a marketing tool, a thing to arouse interest and stand out from the crowd.

As the last group, Bitcoin exchange executives also stress fin-tech innovations as the key driver for competitiveness,[4] making Bitcoin an important part of it. From their point of view, merchants should be preparing themselves, as competitive pressures to offer it as a payment solution will rise soon, triggered through increased word-of-mouth between users. Whether Bitcoin can be considered a 'cool trend', there are two different opinions: One manager confirms this, insisting that Bitcoin could become a lifestyle product, but another reminds us that when it comes to money issues, people are by far more conservative when compared to other areas of life.

4.4 Future Potential

The customers' overall opinion about Bitcoin is very positive; they perceive it as very promising. Nevertheless, credit cards, PayPal, and some other payment services are still and will for some time in the future be the most popular online payment method. Customers are aware of the benefits offered by Bitcoin and alternative means of payments in general, but experience a certain level of effort, through a cumbersome registration and administration effort. Issues with credit cards, e.g., manually typing in the card number and security code, seem to be tolerated until better solutions are developed and more transparency is created as to which shops actually accept Bitcoin.

Merchants strongly support cryptocurrency. Regardless of all arguments and restrictions discussed above, from a business perspective, Bitcoin seems to provoke large interest, and is seen as the future (therefore in line with Van Alstyne [33]). Hence, already two of the four interviewees accept Bitcoin. Nevertheless, the current customer demand is very low, so Bitcoin is still a pure niche product and will take time to develop.

As expected, the managers of Bitcoin exchanges have a very optimistic and positive view of Bitcoin's potential. However, the lack of secure storage and usage solutions hinders market penetration. A widespread initiative to push Bitcoin would be welcomed, and the type and extent of regulation will be decisive for its market penetration.

The high importance of educating people and businesses about the actual purpose and functioning of Bitcoin can be regarded as the unifying factor of all groups (in line with Papilloud and Haesler [79]). All interviewees also agreed that the actual timing of widespread adoption cannot be judged at all.

Interestingly, interviewees of all groups also mentioned the underlying blockchain as an idea and technology that could have a far-reaching, disruptive potential outside of

[4] European Commission: http://europa.eu/rapid/press-release_MEMO-13-719_en.htm.

the payment area. Applications as diverse as document version control, proof of whether a person's vote has been counted or clear identification for e-government and open data solutions were addressed.

5 Conclusion

5.1 Contributions

With this paper, we have contributed to answering several research questions relating to Bitcoin's relative advantages and disadvantages, the drivers and barriers of adoption, and the perceived future potentials from user and professional perspectives.

First, perceived ease of use among stakeholders is still considered rather low. Issues of usability of mobile wallets, implementation, storage, and transfer as well as offering user training are not solved satisfactorily yet. Risks like volatility, security, and accessibility remain. However, people who are using it confirm its high level of convenience. In total, some homework needs to be done here.

Second, perceived usefulness is confirmed for the main part, albeit not yet for the majority of online shoppers. Low transaction fees are central and pivotal to all interviewed individuals. A smaller influence is given to a somewhat higher anonymity and lower risk of fraud as well as faster claim of payments for merchants. Potential international reach and acceptance were also convincing. Bitcoin as a new investment class is controversial, but the high added value for less-developed countries is not.

Third, subjective norm is somewhat split. To use Bitcoin in innovation and in creating competitive advantage is for all groups of high importance. Peer-influence, the fact of being trendy or not, as well as lifestyle issues have less influence on the subjective norm of using Bitcoin.

Fourth, future potential was questioned by comparing it with alternatives. All of them are deemed important, but with a diminishing degree. Current market penetration is judged as very low, recognizing Bitcoin as a niche phenomenon. Subjective future potential is quite the opposite, with all groups assigning a boost in penetration and eventually becoming 'a big thing'. The actual timing of this boost was, however, not answered by anyone.

Fifth, this piece of research has found some evidence that Bitcoin, due to its completely new way of working without a central institution overseeing it, is about to become a serious new player in the online payment market. However, society and businesses are still far from embracing Bitcoin in their daily lives, making educating people about Bitcoin's advantages and use necessary. But Bitcoin's conditions seem good to become fully accepted and trusted by first addressing certain industries and later convincing the masses. One step in that direction may be seen in the latest tech-giant Microsoft accepting Bitcoin for Apps, Windows licenses, Windows Phone, and Xbox in December 2014.[5] Hence, policy-makers and merchants should be prepared.

[5] http://blogs.microsoft.com/firehose/2014/12/11/now-you-can-exchange-bitcoins-to-buy-apps-games-and-more-for-windows-windows-phone-and-xbox.

Finally, the interviews brought additional areas of use of the blockchain scheme to the surface. The possibility to issue unique identifiers to things and people alike opens up immense opportunities in the future.

From a theoretical point of view, we help to expand the body of knowledge about users' acceptance, views, opinions, and feelings about cryptocurrencies—a very new social phenomenon. Additionally, online shops, policy-makers, and regulators should be alarmed that Bitcoin is indeed not just a crazy idea from some IT nerds, but a possible game changer for the future. They should take adequate measures to handle it.

5.2 Limitations and Future Research

Our paper suffers from some limitations. First, only 13 individuals were interviewed, which makes generalizations difficult, as is the case with most inductive, qualitative research. Based on these interview findings, as a follow-up, hypotheses could be formulated that are then tested in a large-scale quantitative survey in a deductive research setting. These survey results would then make generalizations possible. Second, there may be a bias that almost one-third of the interviewees work for Bitcoin exchanges and hence have an apparent interest in seeing cryptocurrencies as too positive. However, very new and potentially disruptive innovations demand to include the voice of professionals, who can judge the whole scope of the research object [65].

Overall, the paper lays a first foundation to spark future research avenues. Comparing developments in different parts of the world could debunk differences due to risk-aversion (e.g., very high in Germany), methods of payment (e.g., widespread credit card use in the US), or technical availability (e.g., lack of traditional banking infrastructure in developing countries). Lastly, including the opinions of people from regulatory and administrative authorities could help in getting the complete picture of cryptocurrencies. And lastly, focusing more on the non-payment aspects of the blockchain technology to revolutionize society in the future may yield great additions to the body of knowledge in the social sciences.

These are just a few ideas contributing to the massive research effort that is still needed to completely understand the emergent and possibly disruptive nature of cryptocurrencies, and to see whether the blockchain can also contribute to a more just society, based on open platforms and open data with access for everyone.

References

1. Foster, G., Gupta, M., Palmer, R.: Business-to-Business Electronic Commerce. Cases in Strategic-Systems Auditing (1999)
2. Bourgeois, R.: Impact of PayPal, Google, Amazon & Emerging Payment Providers on Visa, MasterCard & Payment Industry. The Long-View: 2010 Edition – U.S. Perspectives. Bernstein Global Wealth Management, pp. 195–206 (2010)
3. Olsen, Ø.: Annual Report on Payment Systems 2012 (2013)
4. Koley, T.: End of duopoly in credit card payment scheme industry. J. Econ. Finance **4**, 67–76 (2014)

5. Gonggrijp, S., Geerling, M., Mallekoote, P.: Successful introduction of new payment methods through 'co-opetition. J. Payments Strategy Syst. **7**, 136–149 (2013)
6. Conrad, L.: En garde! banks and paypal will clash. U.S. Banker **117**, 102 (2007)
7. Pogue, D.: The future is plastic. Sci. Am. **312**, 35 (2015)
8. Dempsey, P.: Is it time to take a reality check on the mobile payments market. Eng. Technol. **9**, 20 (2014)
9. Cusumano, M.A.: The Bitcoin ecosystem. Commun. ACM **57**, 22–24 (2014)
10. Brito, J., Castillo, A.: Bitcoin: a primer for policymakers. Policy **29**, 3–12 (2014)
11. Grinberg, R.: Bitcoin: an innovative alternative digital currency. Hastings Sci. Technol. Law J. **4**, 160–208 (2011)
12. Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M.C., Siering, M.: Bitcoin - asset or currency? revealing users' hidden intention. In: European Conference on Information Systems, pp. 1–14 (2014)
13. Bohr, J., Bashir, M.: Who uses Bitcoin? an exploration of the Bitcoin community. In: Twelfth Annual International Conference on Privacy, Security and Trust (PST), pp. 94–101 (2014)
14. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: European Conference on Information Systems, pp. 2206–2217 (2009)
15. Levy, J., Ellis, T.J.: A systems approach to conduct an effective literature review in support of information systems research. Informing Sci. J. **9**, 181–212 (2006)
16. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Q. **26**, 13–23 (2002)
17. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**, 1165–1188 (2012)
18. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008)
19. Scott, B.: Peer-to-Peer Review: The State of Academic Bitcoin Research (2014). http://suitpossum.blogspot.de/2014/12/academic-bitcoin-research.html
20. Böhme, R., Christin, N., Edelman, B., Moore, T.: Bitcoin: economics, technology, and governance. J. Econ. Perspect. **29**, 213–238 (2015)
21. Verbücheln, S.: How Perfect Offline Wallets Can Still Leak Bitcoin Private Keys (2015)
22. Andrychowicz, M., Dziembowski, S., Malinowski, D., Mazurek, L.: Secure multiparty computations on Bitcoin. In: 2014 IEEE Symposium on Security and Privacy (SP), pp. 443–458 (2014)
23. Babaioff, M., Dobzinski, S., Oren, S., Zohar, A.: On Bitcoin and red balloons. In: Faltings, B., Leyton-Brown, K., Ipeirotis, P. (eds.) EC 2012 Proceedings of the 13th ACM Conference on Electronic Commerce, pp. 56–73. ACM, New York (2012)
24. Feld, S., Schönfeld, M., Werner, M.: Traversing Bitcoin's P2P network: insights into the structure of a decentralised currency. Int. J. Comput. Sci. Eng. (forthcoming)
25. Biryukov, A., Khovratovich, D., Pustogarov, I.: Deanonymisation of clients in Bitcoin P2P network. CoRR abs/1405.7418 (2014)
26. Decker, C., Wattenhofer, R.: Information propagation in the Bitcoin network. In: 2013 IEEE Thirteenth International Conference on Peer-to-Peer Computing (P2P), pp. 1–10 (2013)
27. Baek, C., Elbeck, M.A.: Bitcoins as an investment or speculative vehicle? a first look. Appl. Econ. Lett. **22**, 30–34 (2015)
28. Wu, C.Y., Pandey, V.K.: The value of Bitcoin in enhancing the efficiency of an investor's portfolio. J. Financ. Planning **27**, 44–52 (2014)
29. Ali, R., Barrdear, J., Clews, R., Southgate, J.: Innovations in payment technologies and the emergence of digital currencies. Bank Engl. Q. Bull. **54**, 262–275 (2014)

30. Kristoufek, L.: BitCoin meets google trends and wikipedia: quantifying the relationship between phenomena of the internet era. Sci. Rep. **3**, 1–7 (2013)
31. Kondor, D., Csabai, I., Szüle, J., Pósfai, M., Vattay, G.: Inferring the interplay between network structure and market effects in Bitcoin. New J. Phys. **16**, 1–10 (2014)
32. Raiborn, C., Sivitanides, M.: Accounting issues related to Bitcoins. J. Corp. Acc. Finance **26**, 25–34 (2015)
33. van Alstyne, M.: Why Bitcoin has value. Commun. ACM **57**, 30–32 (2014)
34. Holmquist, E.: Bitcoin and the coming revolution in financial transactions. RMA J. **97**, 23–29 (2014)
35. Angel, J.J., McCabe, D.: The ethics of payments: paper, plastic, or Bitcoin? J. Bus. Ethics, 1–9 (2014)
36. Zarifis, A., Efthymiou, L., Cheng, X., Demetriou, S.: Consumer trust in digital currency enabled transactions. In: Abramowicz, W., Kokkinaki, A. (eds.) BIS 2014 Workshops. LNBIP, vol. 183, pp. 241–254. Springer, Heidelberg (2014)
37. Lin, P., Chung, P.-C., Fang, Y.: P2P-iSN: a peer-to-peer architecture for heterogeneous social networks. IEEE Netw. **28**, 56–64 (2014)
38. Maurer, B., Nelms, T.C., Swartz, L.: "When perhaps the real problem is money itself!": the practical materiality of Bitcoin. Soc. Semiot. **23**, 261–277 (2013)
39. Ahamad, S.S., Nair, M., Varghese, B.: A survey on crypto currencies. In: 4th International Conference on Advances in Computer Science, AETACS, pp. 42–48 (2013)
40. Ametrano, F.M.: Hayek Money: The Cryptocurrency Price Stability Solution. SSRN J., 1–54 (2014)
41. Wiatr, M.: Bitcoin as a Modern Financial Instrument (2014)
42. Spenkelink, H.: The adoption process of cryptocurrencies. Identifying factors that influence the adoption of cryptocurrencies from a multiple stakeholder perspective (2014)
43. Vigna, P., Casey, M.J.: The Age of Cryptocurrency. How Bitcoin and Digital Money are Challenging the Global Economic Order. St. Martin's Press, New York (2015)
44. Dinu, A.: The Scarcity of Money: The Case of Cryptocurrencies. Master' Thesis (2014)
45. Papp, J.: A medium of exchange for an internet age: how to regulate Bitcoin for the growth of e-commerce. Pittsburgh J. Technol. Law Policy **15**, 33–56 (2014)
46. Bayern, S.: Dynamic common law and technological change: the classification of Bitcoin. Wash. Lee Law Rev. Online **71**, 22–34 (2014)
47. Citi GPS: Disruptive Innovations II – Ten More Things to Stop and Think About (2014)
48. Andersson, G., Wegdell, A.: Prospects of Bitcoin. An evaluation of its future (2014)
49. Antonopoulos, A.M.: Mastering Bitcoin: Unlocking Digital Cryptocurrencies. O'Reilly & Associates, Sebastopol (2015)
50. Descôteaux, D.: Bitcoin: More Than a Currency, Potential for Innovation. (2014)
51. Dostov, V., Shust, P.: Cryptocurrencies: an unconventional challenge to the AML/CFT regulators? J. Fin. Crime **21**, 249–263 (2014)
52. Evans-Pughe, C., Novikov, A., Vitaliev, V.: To bit or not to bit? Eng. Technol. **9**, 82–85 (2014)
53. Stokes, R.: Virtual money laundering: the case of Bitcoin and the Linden dollar. Inf. Commun. Technol. Law **21**, 221–236 (2012)
54. Bryans, D.: Bitcoin and money laundering: mining for an effective solution. Indiana Law J. **89**, 441–472 (2014)
55. Brezo, F., Bringas, P.G.: Issues and risks associated with cryptocurrencies such as Bitcoin. In: Berntzen, L., Dini, P. (eds.) SOTICS 2012. The Second International Conference on Social Eco-informatics, pp. 20–26. IARIA, USA (2012)
56. Trautman, L.J.: Virtual currencies: Bitcoin & what now after liberty reserve and silk road? Richmond J. Law Technol. **20** (2014)

57. Pagliery, J.: Bitcoin and the Future of Money. Triumph Books, Chicago (2014)
58. Krugman, P.R.: Bitcoin is evil. New York Times 2012 (2013)
59. Cheung, A., Roca, E., Su, J.-J.: Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. Appl. Econ. **47**, 2348–2358 (2015)
60. Ghauri, P.N., Grønhaug, K.: Research Methods in Business Studies. A Practical Guide. Financial Times Prentice Hall, Harlow (2005)
61. Creswell, J.W., Plano Clark, V.L.: Designing and Conducting Mixed Methods Research. Sage, Thousand Oaks (2007)
62. Denzin, N.K., Lincoln, Y.S.: The Sage Handbook of Qualitative Research. Sage, Thousand Oaks (2011)
63. Hair, J.F., Wolfinbarger, M., Money, A.H., Samouel, P., Page, M.J.: Essentials of Business Research Methods. M.E. Sharpe, Armonk (2011)
64. Miles, M.B., Huberman, A.M.: Qualitative Data Analysis. Sage, Thousand Oaks (1994)
65. Eisenhardt, K.M.: Building theories from case study research. Acad. Manag. Rev. **14**, 532–550 (1989)
66. Kvale, S., Brinkmann, S.: InterViews. Learning the Craft of Qualitative Research Interviewing. Sage, Thousand Oaks (2015)
67. Strauss, A., Corbin, J.: Grounded Theory: Grundlagen Qualitativer Sozialforschung (Basics of Qualitative Research). Beltz, Weinheim (1996)
68. Miles, M.B., Huberman, A.: Michael: drawing valid meaning from qualitative data: toward a shared craft. Educ. Researcher **13**, 20–30 (1984)
69. Hesse-Biber, S.N., Leavy, P.: The Practice of Qualitative Research. Sage, Thousand Oaks (2006)
70. Flick, U.: Managing Quality in Qualitative Research. Sage, London (2008)
71. Shenton, A.K.: Strategies for ensuring trustworthiness in qualitative research projects. Educ. Inf. **22**, 63–75 (2004)
72. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**, 319–340 (1989)
73. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: a comparison of two theoretical models. Manag. Sci. **35**, 982–1003 (1989)
74. Venkatesh, V.: Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf. Syst. Res. **11**, 342 (2000)
75. Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag. Sci. **46**, 186 (2000)
76. Silverman, D.: Doing Qualitative Research. Sage, London (2013)
77. Garcia, D., Tessone, C.J., Mavrodiev, P., Perony, N.: The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. J. R. Soc. Interface **11**, 1–28 (2014)
78. Kostakis, V., Giotitsas, C.: The (a)political economy of Bitcoin. TripleC **12**, 431–440 (2014)
79. Papilloud, C., Haesler, A.: Distinktion: the veil of economy: electronic money and the pyramidal structure of socieries. Scand. J. Soc. Theory **15**, 54–68 (2014)

A Systematic Review of Impediments Blocking Internet of Things Adoption by Governments

Paul Brous[1,2(✉)] and Marijn Janssen[1]

[1] Delft University of Technology, Delft, The Netherlands
{P.A.Brous, M.F.W.H.A.Janssen}@tudelft.nl
[2] Rijkswaterstaat, Delft, The Netherlands

Abstract. The Internet of Things (IoT) has high promises and might provide many benefits, yet has been given scant attention in e-government literature. Within the IoT, physical objects, "things", are networked and connected to the Internet. These "things" are able to identify themselves to and communicate with other devices or "things". There are many impediments blocking the adoption of IoT, and there is limited insight in these barriers. In this paper, impediments for the adoption of IoT are investigated by conducting a literature review and carrying out two case studies. The impediments found in literature were confirmed and extended using the case studies. Results show that impediments are interrelated and occur on the strategic, tactical and operational level. For adoption the impediments needs to be addressed in concert. Research on e-governance can benefit from understanding these interrelated impediments.

Keywords: Internet of things · IoT · Adoption · Open data · e-governance · e-government · Smart cities · Impediments · Barriers · Challenges

1 Introduction

The term, the Internet of Things (IoT) refers to the increasing network of physical objects that feature an IP address for internet connectivity, and the communication that occurs between these objects and other Internet-enabled devices and systems [1–3]. The IoT makes it possible to access remote sensor data and to monitor and control the physical world from a distance, allowing many physical objects to act in unison, though means of ambient intelligence [1]. These devices and the communication between these devices can benefit e-government by providing enough quality data to generate the information required to make the right decisions at the right time [3], but in order to achieve this, a variety of impediments need to be overcome. For example, Inductive loops embedded in the road surface are a key technology for traffic detection. An inductive loop is a simple and reliable way to detect the movement of vehicles over a road surface and is extensively used in traffic responsive traffic signal systems to collect traffic data to optimise signal timings accordingly [2]. Such loops provide data on traffic density, flows and speeds for trend analysis as well as providing a key input to real-time traffic models which predict queues or delays. However, installing these loops is costly and the flow of traffic can be obstructed during installation. Another example is that of the application of sensors for the inspection and testing of levees

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 81–94, 2015.
DOI: 10.1007/978-3-319-25013-7_7

(smart levees) in levee management [4]. The sensors embedded in the levees supply a wide range of data. This data is centrally stored and used for the real time visualization of the measurements in a dashboard displaying the sensor results. The data is then directly interpreted for detection and warning systems. These sensors are increasingly being used for the management and monitoring of water barriers, but the technology and the models required to fully analyse the data are still in their infancy and managers are unable to fully trust the system.

The benefits of IoT for governments are known [3] and often emphasized in work. However, less attention has been given to the impediments, or barriers, of IoT, especially with regards to the management and maintenance of large physical infrastructure, have till now not been investigated systematically. Several researchers mention the need for further research in this area [6]. This research explores systematically the potential impediments of the IoT by investigating real world case studies and reviewing state of the art literature.

The methodology used in this research is described in section two. A first overview of IoT impediments will be presented in section three on the basis of state of the art literature. Explorative case studies at the Directorate General of Public Works and Water Management of the Netherlands will be presented in Sect. 4. The Directorate General of Public Works and Water Management of the Netherlands is commonly known within The Netherlands as "Rijkswaterstaat", often abbreviated to "RWS", and is referred to as such within this research. RWS is part of the Dutch Ministry of Infrastructure and the Environment and is responsible for the design, construction, management and maintenance of the main infrastructure facilities in the Netherlands. The results of the literature review and the case studies, and the potential impediments of IoT adoption in e-governance will be discussed in section five. The results show that IoT has a variety of potential impediments at the strategic, tactical and operational levels. Finally conclusions will be drawn in section six.

2 Research Method

We followed two main research steps to determine the potential impediments of IoT for e-governance. First the common impediments of IoT were identified from a rigorous review of literature. The keywords: "Internet of Things" (or "IoT"), "impediments" (or "barriers"), and "e-governance", returned zero hits within the databases Scopus, Web of Science, IEEE explore, and JSTOR. When we replaced the keyword "e-governance" with "governance", we retrieved the same result. The query [all abstract: ("impediments" OR "barriers") AND "internet of things" AND "e-governance"] searching between 2000 and 2015 returned fifty-five hits in Google Scholar. Removing the word "governance" totally from the search string returned more results (67, 0, 13, 6, and 3170 hits respectively).

We found that a great deal of these articles mentioned IoT as being a potential facilitator for achieving the goal of a Smart City based on IoT technology, and some touched on the impediments, barriers or challenges of the implementation of IoT, but most articles relied on anecdotal evidence. Very few articles found were of a general

nature. We then filtered these results and performed a forward and backward search and selected relevant articles based on the criteria that they specifically referred to potential impediments, barriers or challenges with regards to the use or implementation of IoT within potential e-governance applications. The results of the literature lead to a framework within which we developed the case studies.

We used explorative case studies to extend and refine the list from literature the potential impediments of IoT within e-governance applications as the second main research method. Two cases were studied within the context of RWS, which gave the researchers access to subject matter experts and internal documentation for all the cases. This helped ensure the construct validity of the case studies [8]. The cases were selected based on their use of IoT for e-governance purposes – the unit of analysis being programmes within RWS which use and develop IoT for e-governance purposes. The Netherlands is an e-participation leader according to the United Nations e-government survey [9] (2014). This contributes to the validity of the cases as being good representations of e-governance. The cases under study were selected from different domains within RWS in order to ensure diversity and external validity through replication logic [8], in which each case serves as a distinct experiment that stands on its own as an analytic unit. The domains selected were road management and water management respectively.

We studied two separate cases to refine and extend the list of benefits from literature. In the Netherlands there are many similarities, but also, subtle differences in how processes are managed between the "wet" or, water management domain and the "dry", or road management domain. For example, when dealing with objects in the water domain, it is not always possible to be highly accurate with regards to location, as objects placed in water are less static and are often more difficult to physically get to than objects on the ground. We felt it necessary to select cases from both these domains in order to gain a more rounded perspective of the implementation of IoT within e-government in the Netherlands. The cases selected were: 1. Sensor information gathered for the purpose of road management; 2. Sensor information gathered for the purpose of water management. The first case deals with sensor information gathered by RWS with regards to traffic and road management. The second case study deals with sensor information gathered by RWS with regards to water management. The case studies were explorative in method and descriptive in nature. Unstructured interviews were held with managers, subject matter experts, and consultants within RWS. Internal documentation was also studied. Finally, the results of the cases were shared with and verified by subject matter experts within RWS. The pattern-matching technique [11] was used to analyse the case study evidence. Such logic compares an empirically based pattern (findings from the case studies) with a predicted pattern suggested by the literature review, strengthening the internal validity of the research [8]. The technique was applied in the following way. First the common impediments of IoT found in literature were listed. These common impediments were then compared with the evidence of the impediments of IoT from the case study analysis. There were several iterations throughout the research as each case introduced new potential impediments. The potential impediments of the IoT are expressed in italics within this paper.

3 Literature Background

Public and private organizations are increasingly turning to the IoT as new sources of data. However, there are several technological and regulatory challenges that need to be addressed. Scarfo (2014) believes that the most important of them are related to data ownership, security, privacy and sharing of information [12]. It is clear that the implementation of IoT for e-governance faces a variety of impediments. We list the possible impediments of IoT according to strategic/political, tactical and operational divisions. This is a popular division [13, 14], which is suitable for e-governance research.

3.1 Strategic/Political

Skarmeta et al. (2014) consider security and privacy to be the main obstacles for a full acceptance of IoT [15]. IoT devices generate a huge amount of data. The sensitivity levels of the information, is a crucial aspect to be considered by the access control mechanism. Disclosure of user data could reveal sensitive information such as personal habits or personal financial information. The unauthorized access to this information can severely impact user privacy [12, 15–20].

Data produced by IoT devices can be combined, processed and analysed, creating additional insights, so it is important to allow access to data generated by other IoT devices, whilst preventing the unauthorized access and misuse of this information [15]. However, as the IoT becomes more widespread, new security issues become evident [21]. Ortiz et al. (2013) believe that whilst these technologies have been widely investigated for traditional technologies such as relational databases, so far there are no convincing solutions for providing fine grained access control. This hinders the uptake of IoT in e-governance applications dealing with sensitive data [12, 15–22]. In this way, IoT requires novel approaches to ensure the safe and ethical use of the generated data [23], requiring a strong data governance [12, 18, 19, 24, 25]. A weak form of data governance can impede the safe and ethical use of data generated by IoT devices.

A lack of, or poorly coordinated, policies and regulations regarding IoT can also greatly impede the implementation and application of IoT. According to Misuraca (2009), IoT brings with it a wealth of new business opportunities. There is enormous scope for developing applications and selling new services [26]. Governments need to develop policy and regulations and position themselves carefully within this arena [19, 22, 25]. In this regard, public organisations should consider carefully the role they play in the enabling IoT development in the private sector. Market forces of supply and demand can play substantial roles in the success or failure of IoT [17, 26–28]. For example, according to Qiao et al. (2012) the IoT industry, in the short term, will demonstrate an inevitable outbreak growth at the growth stage of the Industry Lifecyle Theory [29]. The internal mechanism of explosive growth is that the whole networking industry chain achieves linkage development between supply and demand [27], but there is a danger that IoT may miss this linkage development due the chain of IoT industry being blocked by a tactical barriers such as a lack of technology breakthroughs, standards bottlenecks and cost barriers [27].

3.2 Tactical

Although reduction in overall costs is an often cited benefit of IoT for e-governance [3], many researchers also cite high development and implementation costs as an important impediment to the implementation and application of IoT for e-governance [17, 19, 22, 27, 30]. According to Yazici (2014), high maintenance costs are often rated as the largest impediments to IoT implementation. A fully functional IoT system based on RFID technology can be substantial. By way of example, Yazici (2014) quotes Wal-Mart's vendors as having spent US$1 to US$3 million on a RFID implementation.

The Internet of things is more than one device, application or network. In order to ensure sustainable connectivity, all interfaces and communication protocols require unified industry standards [17]. However, Fan et al. (2014) believes that the large number of standards-setting organizations has led to a situation in which the top standard has not yet been set. Vendors are free to choose which standard they find best fits their production line, leading to a wide variety of available types which impedes interoperability and integration of data [12, 17, 22, 25, 28, 31]. According to Zeng et al. (2011), there are two methods to integrate things to the Internet: direct integration and indirect integration.

Home appliances are usually directly integrated whilst RFIDs are indirectly integrated through a RFID reader with an embedded server. It is not uncommon for a system to utilise both methods. But IoT requires that a large number of devices be integrated with the existing Internet. These devices can be diverse in terms of data communication methods and capabilities, computational and storage power, energy availability, adaptability, mobility, etc. The heterogeneity at the device level is, in this way, a serious impediment to IoT adoption [16]. This is especially complex as consumers of data are also heterogeneous [16]. Their needs vary in terms of capabilities and data quality. Furthermore, different applications might implement disparate data processing or filtering [16]. Zeng et al. (2011) believe that it is these heterogeneity traits of the overall system that make the design of a unifying framework and the communication protocols a very challenging task, especially with devices with different levels of capabilities. This issue is exacerbated in a large distributed environment.

According to Zeng et al. (2011), Universal Plug and Play (UPnP) is currently the most popular solution for personal network implementation [16]. However, there is no authentication protocol proposed for UPnP. All devices are allowed to configure the other devices on the personal network, without any user control. This can result in a critical security issue when the smart things become available on the Internet. The attention given to security by a number of authors [12, 15–18, 22] suggests that a lack of security standards is becoming a serious impediment to IoT implementation. Whilst there are many standard technologies and protocols to address many security threats, the severe constraints on the IoT devices and networks prevent a straightforward implementation of these solutions [15]. Furthermore, IoT devices generally have to work in harsh, uncontrolled environments, where they may be prone to attacks, misuse or malicious intentions [15]. If a mission critical system is hacked or becomes unavailable, this can lead to a breakdown of trust in the system [15, 17].

According to Kranenburg et al. (2014), the success of user-centric services based on IoT technology depends primarily on people participating and sharing the information

flows [20]. Willingness on the part of people to participate in these systems is therefore required [16, 17, 19, 21, 24, 28, 30, 32]. Kranenburg et al. (2014) believe that this willingness is predominantly dependent on the perception of people: the perceived trust and confidence in IoT and the perceived value that the IoT generates for them. The greater the trust of users in the IoT, the greater their confidence in the system and the more willing they will be to participate [20]. A lack of trust in the system can be a strong impediment to the effectiveness of IoT.

3.3 Operational

Operational barriers include human capital issues such as difficulty in employing qualified personnel, lack of specialists, and personnel skill shortage to operate new applications [19, 22, 32], as well as insufficient IoT oriented training and educational activities [22]. Harris et al. (2015) also identify personnel reluctance to change or to learn new technology as a barrier. A lack of understanding about how IoT works, the possible benefits, and how to make the business case for IoT implementation were also found to be barriers by a number of researchers [19, 32–34]. Reyes et al. (2012) also includes calculating the return on investment and the payback period in this category [34].

Operational barriers also include technical issues such as limitations in information technology (IT) infrastructural capabilities [12, 16–20, 28, 35]. According to Scarfo (2014), the main technological challenges include architecture, energy efficiency, security, protocols and quality of service [12]. An important enabler for the IoT is to permit others to access and use the things that have been published publicly on the internet. It should be possible for users to make use of things that others have shared and to make use of things in their own applications, perhaps in ways unanticipated by the owner of the thing [31]. This requirement means we need a sophisticated set of mechanisms to publish and share things and ways to find and access those things [31]. A lack of these mechanisms as well as the level of knowledge required to implement, manage and maintain the available toolsets can form an important barrier to implementing IoT for e-governance purposes.

Data management issues are also of concern. Public organisations are often faced with a complex legacy of data and applications when implementing IoT solutions [24]. Many public organisations may have several generations of systems running in parallel, and much of the data fed into the system has been done manually, with associated risks in terms of data quality [24, 25, 31].

In short, IoT faces a variety of barriers related to the proper use (privacy and security for example) and proper management of the data collected by the vast number of interconnected things. *Strategic/political barriers are: data privacy issues, data security issues, weak or uncoordinated data policies, weak or uncoordinated data governance, and conflicting market forces. Tactical barriers include: costs, interoperability and integration issues, acceptance of IoT, and trust related issues. Operational issues are: a lack of sufficient knowledge regarding IoT, IT infrastructural limitations, and data management issues.*

4 Case Studies

The goal of the case study research was to refine and extend the list of impediments from literature and to understand the real life impediments of IoT in the most complete way possible. The case study research therefore involved the use of multiple data collection methods. The cases were selected from the primary processes of RWS. Generally IoT is implemented in RWS with the specific intention of ensuring the good working of the primary processes to achieve the primary objectives. In RWS there is a subtle divide in how processes are managed between the water management domain and the road management domain. In order to gain a rounded perspective of the benefits of IoT within RWS, it was believed necessary to select cases from both these domains.

4.1 Case Study 1: Road Management Data Collection at RWS

RWS builds, manages and maintains the Dutch national highways. Correct data is required to do this effectively. Over the years, RWS has developed several methods for obtaining the necessary data from the highways it manages, collecting, processing and making the data available to traffic and road management teams. Measurements are generally made by placing sensors in the road in many different locations. These sensors produce large amounts of data which is mainly used in mid-term planning, long term projections, air quality predictions and noise calculations which have an impact on health and safety measures as well as the environmental impact, and improving service efficiency with regards to road works management.

RWS has created a national network of monitoring points, the "Weigh in Motion" (WIM) network. At present, RWS estimates that at least 15 percent of freight traffic on the Dutch national road network is overloaded. Overloading of heavy vehicles causes road pavement structural distress and a reduced service lifetime [36]. Effectively reducing overloading reduces the damage to the road infrastructure, lengthening the road's lifetime and reduces the frequency of maintenance. The WIM network, consisting of measuring stations in the road on which the axle loads of heavy traffic is weighed, is used to support the enforcement of overloading by helping the enforcement agency to select overloaded trucks for weighing in a static location. The WIM system is one of the most advanced measuring systems in the world. Between 2010 and 2013, RWS built a nationwide WIM network with a total of 18 measuring points. The network consists of 6 newly remodelled measuring stations and 12 new measuring stations. The network provides access to the actual load of the main road, about peak times when it comes to overcharging and it provides RWS with the ability to collect information concerning the compliance behaviour of individual carriers. This forms the basis for business inspections and legal follow-up programs.

RWS faces and has faced a variety of impediments and challenges during the implementation and maintenance of the WIM network. There are different perceptions of the level of ambition pursued by the WIM maintenance process. For example, According to RWS officials, RWS has not yet implemented a structured learning cycle with regards to data quality – "the quality of the data has not been quantified, and

solving data quality issues is incident driven". In this regard, learning takes place in practice and is not formally addressed. Although there have been no direct accusations made between departments, there is also little inter-departmental trust exhibited. According to RWS officials, "Implementation of new technology takes too long, and the implementation process is difficult to follow". There appears to be insufficient knowledge and expertise within the CID to independently manage the WIM systems [38]. The CID reports only on the technical availability of the systems and no information can be provided regarding the performance of the WIM network. RWS is unable to guarantee the reliability of the data due to a lack of a framework of standards. Requirements that exist for managing WIM are not included in the project tender. In 2011, the management of the database with the WIM data was transferred to the Inspectorate General for the Environment and Transport (IET). However the related expertise was not successfully shared. At the present moment, RWS has no access to the data held in the IET databases. The technical requirements were incomplete and some still need to be developed. IET systems are not yet ready to automatically manage the data. There are several legacy issues as technical management is only focused on the availability of the current IT systems. At the time of writing there were technical problems with the license plate recognition system. Governance and mandate appears unclear – it is unclear how the process is coordinated, and the IT supply organisation, the Central Information Department (CID), is unaware that they are also responsible for the sharing of data within the WIM systems. There is no single substantive authority that brings the parties from the entire supply chain together (there is no single authority that assumes responsibility for the entire chain). There is no well-designed change process; changes in the maintenance process are difficult to implement. The representative of the CID in the steering committee has no mandate.

The impediments for the adoption of IoT by e-governance identified in this case are: *1. Strategic/political barriers: data privacy issues, weak or uncoordinated data policies, weak or uncoordinated data governance, and conflicting market forces. Tactical barriers include: interoperability and integration issues, acceptance of IoT, and trust related issues. Operational issues are: a lack of sufficient knowledge regarding IoT, IT infrastructural limitations, and data management issues.*

4.2 Case Study 2: Water Management Data Collection at RWS

Information regarding water quantity and water quality is essential for the primary processes of Rijkswaterstaat: ensuring that flooding does not occur, sufficient clean water and smooth and safe traffic on the water. RWS also collects data on biology and chemistry, measuring nutrients and (micro) pollutants in surface water, suspended matter, sediment and aquatic animals.

The National Water Measurement Network, at RWS known as "Landelijk Meetnet Water" (LMW), is a facility that is responsible for the acquisition, storage and distribution of data for water resources. LMW has more than 400 data collection points using a nationwide system of sensors. The data is then processed and stored in the data centre and is made available to a variety of systems and users. The LMW was created from the merger of three previous existing monitoring networks: the Water Monitoring

Network, which monitored inland waterways such as canals and rivers; the Monitoring Network North, which monitored North Sea oil platforms and channels; and the Zeeland Tidal Waters Monitoring Network which monitored the Zeeland delta waterways. Four main types of measurement activities can be identified: water quantity, water quality, meteorological data and control information on infrastructure. The LMW measures a wide variety of hydrological data such as water levels, flow rates, wave heights and directions, flow velocity and direction, and water temperature. The LMW also measures meteorological data such as wind speed and direction, air temperature and humidity and air pressure amongst others. This meteorological data is collected in close collaboration with the Dutch Royal Meteorological Institute. The LMW provides a complete technical infrastructure for gathering and distribution of data and delivers the data to various stakeholders within and outside RWS.

According to RWS officials, there is often not clear who is ultimately responsible for the entire information chain. It has been discovered that the different departments work with different targets [39]. For example, the goals of the project organisation are to get production legitimacy of payment and to execute system contract management whilst the goals of the asset manager are to prevent flooding, to show demonstrable compliance with the statutory requirements and to reduce probability of failure. This results in different levels of the organization addressing different points of discussion. RWS officials report that a major impediment is "the lack of agreement and decision-making regarding vital strategic choices such as the form of contract management, the tender strategy, and outsourcing of personnel". The experience is that there is a failure to align the implementation with business targets as management teams tend to focus purely on the supply chain with a lack of awareness for possible risks or opportunities that occur outside of the supply chain. Whilst agreements occur between management teams, collaboration does not always occur in operations. There appear to be significant impediments regarding the coordination of activities and projects. RWS officials also quoted a lack of trust in the private sector to be able to manage and maintain the systems adequately. The perspective of RWS is that the private sector is not adequately developed regarding the necessary technical knowledge required by such an intricate system. RWS officials have stated that one of the reasons for this perspective is the intricacy of the RWS technical architecture itself. The experience is that the current architecture and legacy data make future integration of data very difficult.

The impediments for the adoption of IoT by e-governance identified in this case are: *1. Strategic/political barriers: data security issues, weak or uncoordinated data policies, weak or uncoordinated data governance, and conflicting market forces. Tactical barriers: interoperability and integration issues, acceptance of IoT, and trust related issues. Operational issues: a lack of sufficient knowledge regarding IoT.*

5 Discussion

The objective of this research was to identify potential impediments of the IoT for e-governance purposes. The IoT is important because a physical (or sensor) object that is able to communicate digitally is able to relate not only to a single entity, but also

becomes connected to surrounding objects and data infrastructures. This allows for a situation in which many physical objects are able to act in unison, by means of ambient intelligence [1]. These devices and the communication between these devices can benefit e-government by providing enough quality data to generate the information required by government and citizens to make the right decisions at the right time.

We used two main research methods: (1) a literature review, (2) analysis of two IoT case studies. The literature review provided us with an overview of the existing body of knowledge, allowing us to analyse where gaps in knowledge or focus occur. It also provided definitions for the key concepts and helped develop a broader knowledge base in the research area. Case study research is a widely used qualitative research method in information systems research, and is well suited to understanding the interactions between information technology-related innovations and organizational contexts [40]. Following the advice of Yin (2003), the protocol used in the case study included a variety of data collection instruments. In order to counter the possible influences of bias, multiple research instruments were employed to ensure construct validity through triangulation [8].

The results of the literature review and the case studies demonstrate that the IoT is faced with a variety of impediments with regards to adoption b which correlates with impediments identified in the literature review. Table 1 below lists the main impediments of IoT, differentiating between strategic, tactical and operational benefits.

Formulating strategy requires defining goals and initiatives based on available resources and an assessment of the internal and external environments in which the organization competes [41]. Strategic impediments can therefore exert an important influence on an organization's likelihood of success. IoT is capable of providing a continuous stream of "trusted" data which managers can use to make informed, decisions, but the adoption of IoT for this purpose needs to be carefully coordinated by strong data policies and strong governance of the data with a purposeful awareness of opposing market forces and the capability of the private sector to provide critical services. Public sector organisations also need to address data privacy and data security issues for IoT adoption to be successful. These issues are interrelated as legal frameworks and strong policies provide guidelines within which organisations can face the pitfalls placed by security and privacy issues.

The main impediments appear to manifest during implementation, once organisations decide to operationalize the business plan. At that stage it becomes clear that although the technology is ready for widespread implementation, IoT remains an innovation which needs to not only integrate with current legacy systems but for which standards, policy, and legal frameworks still need to be developed with regards to social, technical and ethical issues.

Achieving a strategic plan or objective requires the administrative process of selecting among appropriate ways and means. Tactical planning is short range planning that emphasizes the current operations of various parts of the organization [42]. The case studies show that a good deal of attention should be paid to coordinating "soft", organisational issues such as trust and acceptance of IoT solutions as well as to the coordination of harder, technical, issues which require standardization in order to ensure interoperability and integration of data and systems. Significantly, costs were

Table 1. Impediments of IoT for e-governance in relation to the case studies.

		Impediments	Literature	Case 1	Case 2
Strategic	Social Responsibility	Data privacy issues	✓	✓	
		Data security issues	✓		✓
		Lack of legal framework	✓	✓	✓
	Productivity	Weak or uncoordinated data policies	✓	✓	✓
		Weak or uncoordinated data governance	✓	✓	✓
	Market standing	Conflicting market forces	✓	✓	✓
Tactical	Profitability	Costs	✓		
	Physical resources	Interoperability and integration issues	✓	✓	✓
		Lack of a framework of standards	✓	✓	✓
	Worker attitude	Acceptance of IoT	✓	✓	✓
		Trust related issues	✓	✓	✓
Operational	People	Lack of sufficient capabilities/knowledge	✓	✓	✓
	Technology	IT infrastructural limitations (issues with legacy systems)	✓	✓	
	Processes	Data management issues (data quality issues)	✓	✓	

only mentioned in the case studies as an impediment with regards to a negative business case. Since the primary processes of RWS are directly connected with health, safety and security of Dutch citizens, cost was disregarded as secondary to achieving the primary objective. It is possible that cost may be a more significant impediment in countries with less accessibility to the necessary funding than RWS.

A primary use of IT in government is to improve the efficiency of government operations [43]. As with many other organisations RWS uses IoT as a tool in industrial automation, in which simple manual tasks such as opening and closing bridges are automated. This reduces very low-level coordination work that was previously executed by humans, but complete automation or outsourcing of work can lead to a lack of sufficient knowledge within the organisation regarding the technique and the management of the data. This situation can develop into a significant impediment to the maintenance of IoT in e-governance applications. Technology continues to advance, but whilst many RWS officials were confident that technology was generally not a serious impediment, it is important to ensure that chose technique is compatible with the IoT architecture.

6 Conclusion

This paper represents one of the first papers on IoT for e-government. There has been limited research in the field of e-government regarding IoT, and there is much potential as expressed by the potential benefits, but the adoption of IoT within e-governance applications requires careful preparation and coordination. The IoT makes it possible to access remote sensor data and to monitor and control the physical world from a distance, and combining and analysing captured data allows governments to develop and improve services which cannot be provided by isolated systems, but this can only be achieved by addressing the potential impediments of to IoT adoption at all levels.

This research provides a systematic insight into the potential impediments of the IoT for e-government purposes by means of case study analysis and a review of literature. The research shows that impediments range from the political to the operational level. Specifically impediments for e-government can be attributed to data privacy issues, data security issues, weak or uncoordinated data policies, weak or uncoordinated data governance, and conflicting market forces, costs, interoperability and integration issues, acceptance of IoT, and trust related issues, a lack of sufficient knowledge regarding IoT, IT infrastructural limitations, and data management issues.

Many of the issues are interrelated; interoperability and integration issues have a direct impact on costs and on trust in the systems, and many issues can be resolved with sufficient knowledge and capabilities within the organisation. But the issues do need to be resolved in concert. It is important that governments address dominant impediments, such as privacy and security issues, within public policy and legal frameworks to assist public organisations with implementation of IoT. Similarly, technical and knowledge issues are very much interrelated with a lack of standards and impediments regarding interoperability and integration of data.

Acknowledgements. We acknowledge and thank the people of the Rijkswaterstaat who gave of their time and expertise during the case study research. This research is funded by Rijkswaterstaat.

References

1. Ramos, C., Augusto, J.C., Shapiro, D.: Ambient intelligence-the next step for artificial intelligence. IEEE Intell. Syst. **23**, 15–18 (2008)
2. Hounsell, N.B., Shrestha, B.P., Piao, J., McDonald, M.: Review of urban traffic management and the impacts of new vehicle technologies. IET Intell. Transp. Syst. **3**, 419–428 (2009)
3. Brous, P., Janssen, M.: Advancing E-Government Using the Internet of Things: A Systematic Review of Benefits. Presented at the IFIP Egov (2015)
4. Stichting Ijkdijk: livedijk-utrecht. http://www.ijkdijk.nl/nl/livedijken/livedijk-utrecht
5. RTV Utrecht: Proef met dijksensoren in Woerden en Nieuwegein. http://www.rtvutrecht.nl/nieuws/859256/proef-met-dijksensoren-in-woerden-en-nieuwegein.html
6. Marche, S., McNiven, J.D.: E-Government and E-Governance: the future isn't what it used to be. Can. J. Adm. Sci. Rev. Can. Sci. Adm. **20**, 74–86 (2003)

7. Haller, S., Karnouskos, S., Schroth, C.: The internet of things in an enterprise context. In: Domingue, J., Fensel, D., Traverso, P. (eds.) FIS 2008. LNCS, vol. 5468, pp. 14–28. Springer, Heidelberg (2009)
8. Yin, R.K.: Case Study Research: Design and Methods. SAGE Publications, Chicago (2003)
9. United Nations, Department of Economic and Social Affairs: United Nations e-government survey 2014: e -government for the future we want. United Nations, New York (2014)
10. Eisenhardt, K.M.: Building Theories from Case Study Research. Acad. Manage. Rev. **14**, 532–550 (1989)
11. Trochim, W.M.: Outcome pattern matching and program theory. Eval. Program Plann. **12**, 355–366 (1989)
12. Scarfo, A.: Internet of things, the smart X enabler. In: 2014 International Conference on Intelligent Networking and Collaborative Systems (INCoS), pp. 569–574 (2014)
13. Ivanov, D.: An adaptive framework for aligning (re)planning decisions on supply chain strategy, design, tactics, and operations. Int. J. Prod. Res. **48**, 3999–4017 (2010)
14. Ackoff, R.L.: Towards a system of systems concepts. Manag. Sci. **17**, 661–671 (1971)
15. Skarmeta, A.F., Hernandez-Ramos, J.L., Moreno, M.V.: A decentralized approach for security and privacy challenges in the internet of things. Presented at the 2014 IEEE World Forum on Internet of Things, WF-IoT 2014 (2014)
16. Zeng, D., Guo, S., Cheng, Z.: The web of things: a survey. J. Commun. **6**, 424–438 (2011)
17. Fan, P.F., Wang, L.L., Zhang, S.Y., Lin, T.T.: The research on the internet of things industry chain for barriers and solutions (2014)
18. Hummen, R., Henze, M., Catrein, D., Wehrle, K.: A cloud design for user-controlled storage and processing of sensor data. Presented at the CloudCom 2012 - Proceedings: 2012 4th IEEE International Conference on Cloud Computing Technology and Science (2012)
19. Yazici, H.J.: An exploratory analysis of hospital perspectives on real time information requirements and perceived benefits of RFID technology for future adoption. Int. J. Inf. Manag. **34**, 603–621 (2014)
20. van Kranenburg, R., Stembert, N., Moreno, M., Skarmeta, A.F., López, C., Elicegui, I., Sánchez, L.: Co-creation as the Key to a public, thriving, inclusive and meaningful EU IoT. In: Hervás, R., Lee, S., Nugent, C., Bravo, J. (eds.) UCAmI 2014. LNCS, vol. 8867, pp. 396–403. Springer, Heidelberg (2014)
21. Ortiz, P., Lazaro, O., Uriarte, M., Carnerero, M.: Enhanced multi-domain access control for secure mobile collaboration through linked data cloud in manufacturing. In: IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013, pp. 1–9 (2013)
22. Harris, I., Wang, Y., Wang, H.: ICT in multimodal transport and technological trends: Unleashing potential for the future. Int. J. Prod. Econ. **159**, 88–103 (2015)
23. Roman, R., Najera, P., Lopez, J.: Securing the internet of things. Computer **44**, 51–58 (2011)
24. Gilman, H., Nordtvedt, J.-E.: Intelligent energy: the past, the present, and the future. SPE Econ. Manag. **6**, 185–190 (2014)
25. Stephan, E.G., Elsethagen, T.O., Wynne, A.S., Sivaraman, C., Macduff, M.C., Berg, L.K., Shaw, W.J.: A linked fusion of things, services, and data to support a collaborative data management facility. In: 2013 9th International Conference Conference on Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), pp. 579–584 (2013)
26. Misuraca, G.: Futuring e-Government: governance and policy implications for designing an ICT-enabled knowledge society. In: Proceedings of the 3rd International Conference on Theory and Practice of Electronic Governance, pp. 83–90. ACM, New York (2009)
27. Qiao, H., Wang, G.: An analysis of the evolution in Internet of Things industry based on industry life cycle theory (2012)

28. Wiechert, T.J.P., Thiesse, F., Michahelles, F., Schmitt, P., Fleisch, E.: Connecting mobile phones to the internet of things: a discussion of compatibility issues between EPC and NFC. Presented at the Association for Information Systems - 13th Americas Conference on Information Systems, AMCIS 2007: Reaching New Heights (2007)
29. Audretsch, D.B., Feldman, M.P.: Innovative clusters and the industry life cycle. Rev. Ind. Organ. **11**, 253–273 (1996)
30. Nam, T., Pardo, T.A.: The changing face of a city government: a case study of Philly311. Gov. Inf. Q. **31**(Supplement 1), S1–S9 (2014)
31. Blackstock, M., Lea, R.: IoT mashups with the WoTKit. In: 2012 3rd International Conference on the Internet of Things (IOT), pp. 159–166 (2012)
32. Speed, C., Shingleton, D.: An Internet of cars: Connecting the flow of things to people, artefacts, environments and businesses. Presented at the Sense Transport 2012 - Proceedings of the 6th ACM Workshop on Next Generation Mobile Computing for Dynamic Personalised Travel Planning (2012)
33. Pedro, M.: Reyes, patrick jaska: is RFID right for your organization or application? Manag. Res. News. **30**, 570–580 (2007)
34. Reyes, P.M., Li, S., Visich, J.K.: Accessing antecedents and outcomes of RFID implementation in health care. Int. J. Prod. Econ. **136**, 137–150 (2012)
35. Prasad, K.H., Faruquie, T.A., Joshi, S., Chaturvedi, S., Subramaniam, L.V., Mohania, M.: Data cleansing techniques for large enterprise datasets. Presented at the SRII Global Conference, San Jose, USA, April 29 (2011)
36. Mulyun, A., Parikesit, D., Antameng, M., Rahim, R.: Analysis of loss cost of road pavement distress due to overloading freight transportation. J. East. Asia Soc. Transp. Stud. **8**, 1020–1035 (2010)
37. Bagui, S., Das, A., Bapanapalli, C.: Controlling vehicle overloading in BOT projects. Procedia Soc. Behav. Sci. **104**, 962–971 (2013)
38. Evert, H., van der Valk, S.W.: A3 Weigh in Motion (2013)
39. Hopman, F., Meijer, E., Cyt, A., Piarelal, S., Hoogervorst, O.: RWS LMW HHS Adviesrapport. docx
40. Janssen, M.: Designing electronic intermediaries: An agent-based approach for designing interorganizational coordination mechanisms. Doctoral Dissertation, Delft University of Technology, Delft, The Netherlands (2001)
41. Nag, R., Hambrick, D.C., Chen, M.-J.: What is strategic management, really? Inductive derivation of a consensus definition of the field. Strateg. Manag. J. **28**, 935–955 (2007)
42. Tactical planning vs strategic planning. https://managementinnovations.wordpress.com/2008/12/10/tactical-planning-vs-strategic-planning/
43. Castro, D.: Digital Quality of Life: Government. Available SSRN 1285002 (2008)

Understanding the Adoption of Mobile Internet in the Saudi Arabian Context: Results from a Descriptive Analysis

Abdullah Baabdullah[1], Yogesh Dwivedi[1(✉)], Michael Williams[1], and Prabhat Kumar[2]

[1] Management and Systems Section (MaSS), School of Management, Swansea University, Swansea SA2 8PP, UK
{685177,y.k.dwivedi,m.d.williams}@swansea.ac.uk
[2] Department of Computer Science and Engineering, National Institute of Technology, Patna, India
prabhat@nitp.ac.in

Abstract. Utilising Mobile Internet (M-Internet) services would increase socio-economic benefits. Hence, it is necessary to consider the factors that may increase the adoption of M-Internet services within the context of Saudi Arabia. This research aims to examine potential users' intentions towards different variables that may be significant for supporting higher usage of M-Internet services in the domain of the Kingdom of Saudi Arabia. This study embraces the following variables: perceived risk, innovativeness; performance expectancy, effort expectancy, social influence, facilitating conditions, perceived value, hedonic motivation and behavioural intention. Data was collected by means of a questionnaire on a convenience sample that consisted of 600 subjects with a response rate of 69.5 %. The findings gathered from a descriptive analysis suggested that the related variables are perceived as significant by participants and they have strong behavioural intention to use M-Internet services.

Keywords: Saudi Arabia · M-Internet · UTAUT2 · Perceived risk · Innovativeness

1 Introduction

The importance of studying M-Internet within the context of Saudi Arabia is derived from the statistics that the Kingdom has experienced considerable growth of marketplaces in Information and Communication Technology (ICT) in the Middle East; predominantly in the retail and financial services' sectors as well as domestic broadband penetration (STC Group, 2011). For example, between 2005 and 2011, domestic broadband penetration had risen from zero to 44 % and M-Internet penetration had increased from 60 % to 191 % over the same period [26]. The statistics of using various types of M-Internet services in Saudi Arabia refers to the growth in the adoption of this new technology in areas such as M-Advertising. For example, in 2010, it reached a compound annual growth of 31 % according to Analysys Mason [6]. Also, the number of M-Internet users in Saudi Arabia had reached 11.5 million at the end of 2011

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 95–106, 2015.
DOI: 10.1007/978-3-319-25013-7_8

representing 40.5 % of the population [16]. However, the adoption of other types of M-Internet in Saudi Arabia is still in its early stage. For example, only 39 % of M-Internet users in Saudi Arabia used this service for commercial purposes in 2011 [6]. Indeed, enhancing the level of behavioural intention toward using M-Internet in Saudi Arabia is a challenging task especially in the business arena due to the strong presence of cultural and technological parameters [6]. Tackling the factors that might predict behaviour intention towards using M-Internet among potential users will have a positive impact on raising the rate of using M-Internet across the Kingdom. This research will consider performance expectancy, effort expectancy, social influence, facilitating conditions, perceived value, hedonic motivation, perceived risk and innovativeness as probable predictors of behavioural intention amongst potential adopters of M-Internet in the domain of Saudi Arabia. By doing this research, this study will pinpoint the factors that can predict behaviour intention of M-Internet and thereby enhancing the use of M-Internet in the Saudi Arabian context. This paper is structured as follows: section two mentions an overview of M-Internet literature; then the researchers will give an overview of the theoretical basis in the third section. Next, methodology, results and discussion are outlined in the fourth, fifth and sixth sections respectively. In section seven, this study gives a synopsis.

2 Literature Review

Defining M-Internet is quite important before having any theoretical or empirical discussion as there are a lot of mobile strands that have different characteristics [10]. M-Internet can be defined as the "access to the Internet with devices that offer wireless connectivity" [18]. M-Internet relies on the three interconnected elements: the mobile device, a mobile network, and mobile content [18]. Consequently, the M-Internet makes a world where everyone can be connected regardless of time and space [19]. These distinctive traits, and their link to M-Internet, have generated an extraordinary level of M-Internet adoption worldwide. Indeed, there are plenty of studies that have focused on studying M-Internet (e.g. [7, 19, 21]. Within the context of Saudi Arabia, Alwahaishi and Snášel [4, 5] provided their new framework to pinpoint variables that might affect the acceptance of M-Internet and use in Saudi Arabia. Nonetheless, M-Internet reluctance can be attributed to different factors such as the lack of widespread use of mobile phones [32] and inaccessibility of the technological support [12]. However, within the context of Saudi Arabia and from 2004, use of the M-Internet has grown due to the development in wireless technology which led to a number of advantages such as discarding the need for landlines; minimising repetitive software installations; and solving the problem of temporary loss of high-speed Internet in some areas through using mobile devices [6]. Nevertheless, this advancement of M-Internet in Saudi Arabia has recorded a lower level compared with neighbouring countries such as Kuwait, UAE and Qatar. As such, a study was held in 2010 about the broadband subscriptions by using M-Internet, and found that mobile subscriptions in Qatar, UAE, Kuwait, Saudi Arabia were 72 %, 68 %, 65 % and 42 % respectively [9]. The adoption of M-Internet in Saudi Arabia still lags behind neighbouring states for many reasons such as network problems; that is to say, wireless access is limited in remote areas in

Saudi Arabia. Furthermore, even if there was access, frequent network problems have been recorded. According to a study conducted by Analysys Mason [9] in Saudi Arabia, it showed that there is a need to "harmonise the mobile broadband spectrum" in the "700/800 and 2.6 GHz bands for the use by mobile operators." As M-Broadband refers to Internet access (wireless communications) that has taken the place of a mobile phone, this shortage impacts negatively on the mobile broadband services and eventually on M-Internet services [9]. Although the adoption of M-Internet services has experienced a permanent growth in the Saudi context, the magnitude of this growth is still modest when compared with other states. According to the STC Group [26] regarding individual/household purchasing on the Internet in 2009 by either using the M-Internet or a fixed Internet, Saudi Arabia was way behind the UK and Norway with 2 %, 66 %, and 70 % respectively; Saudi Arabia are also behind Italy (12 %) and Greece (10 %). Interestingly, this modest adoption of M-Internet amongst Saudi citizens was in contrast with the huge increase in the budget of the Saudi telecom sector. As such, in 2009, the per capita telecom investment in Saudi telecom exceeded US $400 m compared with US$150 m in the UK. This information is important to notice that although the telecom revolution industry which is the basis for any activity through M-Internet, the actual use of M-Internet services in Saudi Arabia was way behind the UK. These evidences suggest that M-Internet services have not been heavily used yet, and Saudi users have not yet perceived these services. So, this study will examine the factors that may affect the adoption of this technology by the potential users through examining the factors that might influence behaviour intention of the potential users towards using M-Internet in the Saudi context. The outcomes of this current paper will be as follows: this research will consider UTAUT2 factors as well as perceived risk (PR) and innovativeness (INN) to pinpoint accurately the variables that change the behavioural intention of Saudi potential users in order to use the M-Internet.

3 Theoretical Basis

The Unified Theory of Acceptance and Use of Technology (UTAUT) proposes to predict adopters' intentions to adopt IT. It embraces four variables: performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC). In this theory, PE, EE, and SI are direct determinants of behavioural intention (BI) and actual usage, and FC is a direct determinant of actual usage. Venkatesh et al. [30] confirmed that UTAUT predicted 70 % and 50 % of the variances in BI and usage respectively. Furthermore, a considerable number of writers have applied UTAUT in order to predict BI and usage in the field of M-Internet (e.g. [4, 5, 21, 32]. The UTAUT has been internationally tested, modified and developed over the years in order to explain users' behaviour towards the adoption of the M-Internet, e.g. Venkatesh et al. [29] adopted UTAUT and UTAUT2. Venkatesh et al. [29] developed UTAUT2 for investigating users' adoption of the M-Internet by extending UTAUT to UTAUT2 to bridge the void that occurred in the users' adoption domain. UTAUT2 adds new constructs: i.e. hedonic motivation (HM); price value (PV); and habit (HT) in addition to the UTAUT variables including PE, EE, SI, FC, BI and usage. The addition of these new variants gives more capability of estimation of the variables' intention to

use compared with the UTAUT model. Moreover, in UTAUT2, association between FC and BI has been applied in order to offer a method to test the significance of FC in effecting BI of potential adopters in this study rather than testing the influence of FC over usage as applied in the UTAUT model. Thus, this paper holds UTAUT2 variables for examining the use of the M-Internet by potential adopters in the Saudi Arabian domain. UTAUT2 assembled factors of other models in order to offer a higher level of predictability about the behaviour of potential adopters when using the M-Internet. In detail, PE and EE in UTAUT2 replaced perceived usefulness and perceived ease of use

Table 1. Description of factors

Factor examined	Definition	Example citations that have tested this construct for examining the adoption of mobile applications
Performance Expectancy (PE)	"The degree to which using a technology will provide benefits to consumers in performing certain activities" [28, p. 159]	[29, 30]
Effort Expectancy (EE)	"The degree of ease associated with consumers' use of technology" [28, p. 159]	[3, 28]
Facilitating Conditions (FC)	"Refer to consumers' perceptions of the resources and support available to perform a behaviour" [28, p. 159]	[29, 34]
Social Influence (SI)	"The extent to which consumers perceive that important others (e.g. family and friends) believe they should use a particular technology" [28, p. 159]	[20, 29]
Hedonic Motivation (HM)	"The fun or pleasure derived from using technology and it has been shown to play an important role in determining technology acceptance and use" [9, p. 402]	[27, 29]
Price Value (PV)	"Consumers' cognitive trade-off between the perceived benefits of the applications and the monetary cost for using them" [28, p. 161]	[22, 29]
Innovativeness (INN)	"The willingness of an individual to try out any new information technology" [34, p. 229]	[2, 22]
Perceived Risk (PR)	"The expectation of losses associated with purchase and acts as an inhibitor to purchase behaviour" [17, p. 454]	[24, 34]
Behaviour Intention (BI)	"The intention of the user to use the technology" [33, p. 160]	[29, 33]

Table 2. Scale items of the selected factors

Construct	Item	Source
Performance Expectancy (PE)	PE1. M-Internet will be useful in my daily life	[29]
	PE2. Using M-Internet will help me accomplish things more quickly	
	PE3. Using M-Internet will increase my productivity	
	PE4. Using M-Internet will increase my chances of achieving things that are important to me	
Effort Expectancy (EE)	EE1. Learning how to use M-Internet will be easy for me	
	EE2. My interaction with M-Internet will be clear and understandable	
	EE3. M-Internet will be easy to use	
	EE4. It will be easy for me to become skilful at using M-Internet	
Social Influence (SI)	SI1. People who are important to me think that I should use M-Internet	
	SI2. People who influence my behaviour think that I should use M-Internet	
	SI3. People, whose opinions that I value, prefer that I use M-Internet	
Facilitating Conditions (FC)	FC1. I have the resources necessary to use M-Internet	
	FC2. I have the knowledge necessary to use M-Internet	
	FC3. M-Internet is compatible with other technologies I use	
	FC4. I can get help from others when I have difficulties using M-Internet	
Hedonic Motivation (HM)	HM1. Using M-Internet will be fun	
	HM2. Using M-Internet will be enjoyable	
	HM3. Using M-Internet will be very entertaining	
Price Value (PV)	PV1. M-Internet is reasonably priced	
	PV2. M-Internet is good value for the money	
	PV3. At the current price, M-Internet provides good value	
Perceived Risk (PR)	PR1. Providing M-Internet service with my personal information would involve many unexpected problems	[34]
	PR2. It would be risky to disclose my personal information through this service provider	
	PR3. There would be high potential for loss in disclosing my personal information to this service provider	
	PR4. Using M-Internet services subjects your checking account to financial risk	[17]
	PR5. I think using M-Internet puts my privacy at risk	[33]

(*Continued*)

Table 2. (*Continued*)

Construct	Item	Source
Innovativeness (INN)	INN1. If I heard about M-Internet technology, I would look for ways to experiment with it	[22]
	INN2. Among my peers, I am usually the first to explore new technologies	
	INN3. I like to experiment with new technologies: i.e. M-Internet	
	INN4. In general, I am not hesitant to try out new information technologies	[1]
	INN5. Compared to my friends, I seek out a lot of information about M-Internet services	[2]
	INN6. I would try the new M-Internet service even if, in my circle of friends, nobody has trialled it before	
Behavioural Intention (BI)	BI1. I will use M-Internet in the future	[29]
	BI2. I will always try to use M-Internet in my daily life	
	BI3. I will plan to use M-Internet frequently	

in TAM respectively. FC in UTAUT2 interchanged with perceived behavioural control (PBC) in TPB [11]. Similarly, HM in UTAUT2 replaced the perceived enjoyment variable. Likewise, social impact in UTAUT2 is similar to subjective norm and family influence [29]. Finally, PV stands for price and cost [23]. Venkatesh et al. [29] suggested examining the appropriateness of the UTAUT2 in various contexts. This research adopts the UTAUT2 model factors in addition to two new factors: i.e. perceived risk (PR) and innovativeness (INN) to critically evaluate the impact of these factors over BI of the potential adopters [25]. After reviewing the literature [14, 31], this study will adopt a conceptual model that will consist of candidate variables that might highly affect the adoption of potential users of the M-Internet in the Saudi Arabian context [8]. In order to check these variables, this research followed a quantitative study [15]. Unlike UTAUT2, this study does not consider habit as an independent variable as well as usage as a dependent variable. The reason for not considering habit and usage lies in the fact that this study focuses on the potential user rather than the actual usage of the M-Internet. Hence, it discards actual usage. Furthermore, according to Venkatesh et al. [29], habit has a direct effect on usage; consequently, there is no need for considering the habit variable when not studying usage. The mentioned constructs are viewed in Table 1 as follows:

4 Methodology

According to Table 2, this research tests nine variables: i.e. INN, PR, PE, EE, SI, FC, HM, PV and BI. It refers to the writers to whom this study relies on when considering 35 items. The level of responses was measured by implementing the seven-point Likert scale. Regarding demographic variables, this research embraces gender, age, education,

occupation and monthly income. The researchers employed convenience sampling in three cities: i.e. Jeddah, Riyadh and Dammam.

5 Results

5.1 Response Rate

Table 3 outlines the distributed sample for M-Internet technology, returned responses, incomplete and problematic responses, and finally valid responses.

Table 3. Response Rate

	M-Internet	%
Sample	600	100
Returned responses	436	72.67
Incomplete and problematic responses	19	3.17
Valid responses	417	69.5

5.2 Respondents' Profile and Characteristics

Table 4 below shows thedemographic characteristics of M-Internet respondents:

5.3 Respondents' Profile and Characteristics

According to Table 5, the average mean score and standard deviation were calculated for studied constructs as well as for items of these constructs.

6 Discussion

The descriptive statistics gave mean and standard deviation for each related variable and its items. Generally speaking, reading the average mean and average standard deviation for each variable can be classified into one of these four categories: high mean and high standard deviation, low mean and low standard deviation, high mean and low standard deviation, and low mean and high standard deviation [13]. The related variables, i.e. PE, EE, SI, FC, HM, PV, INN and BI, had a high average mean and a low standard deviation within the accepted limits (±1SD). The high mean, which is combined with the low standard deviation, indicated that the respondents' answers tended to be 'strongly agree' with a high certainty. In contrast, the low mean, which is combined with the low standard deviation, indicated that the respondents' answers tended to be 'strongly disagree' with a high certainty; i.e. perceived risk (PR) as participants seem to be less concerned regarding risks related to M-Internet services. Consequently, it is recommended that future studies should re-examine the significance

Table 4. Respondents' profile and characteristics

Demographic profile	Number of respondents (N = 417)	Percentage (%)
Gender		
Male	247	59.2
Female	170	40.8
Total	417	100
Age		
> = 18–20	55	13.2
21–29	242	58
30–39	68	16.3
40–49	33	7.9
50 and above	19	4.6
Total	417	100
Education		
Less than High School	14	3.4
High School	77	18.5
Diploma	74	17.7
Bachelor	191	45.8
Postgraduate	61	14.6
Total	417	100
Occupation		
Student	53	12.7
Government employee	218	52.3
Private sector employee	97	23.3
Self employed	49	11.7
Total	417	100
Monthly income (Saudi Riyals)		
1000–4000	36	8.6
4001–8000	91	21.8
8001–14000	183	43.9
14001–20000	73	17.5
More than 20000	34	8.2
Total	417	100

level of each one of these variables to obtain a better prediction over the adoption of the M-Internet in the Saudi Arabian context.

6.1 Contribution

Very few studies have been conducted in Saudi Arabia regarding M-Internet. This research has investigated statistically the variables that affect the BI towards adopting the M-Internet amongst potential adopters; thus, this study can suggest the best way to increase the BI of the potential users which is an essential step towards actually

Table 5. Descriptive analysis of measurement items

Construct	Item	Mean	Standard Deviation
Performance expectancy	PE1	6.66	.566
	PE2	6.67	.563
	PE3	6.64	.590
	PE4	6.67	.573
	Average	6.66	.573
Effort expectancy	EE1	6.56	.641
	EE2	6.61	.591
	EE3	6.56	.641
	EE4	6.57	.624
	Average	6.60	.583
Social influence	SI1	6.47	.686
	SI2	6.47	.672
	SI3	6.46	.696
	Average	6.46	.684
Facilitating conditions	FC1	6.45	.671
	FC2	6.46	.646
	FC3	6.45	.678
	FC4	6.20	.885
	Average	6.39	.72
Hedonic motivation	HM1	6.81	.403
	HM2	6.81	.408
	HM3	6.82	.401
	Average	6.81	.404
Price value	PV1	6.01	.732
	PV2	5.97	.751
	PV3	5.96	.770
	Average	5.98	.751
Innovativeness	INN1	6.38	.721
	INN2	5.50	1.312
	INN3	6.42	.743
	INN4	6.52	.676
	INN5	6.19	.953
	INN6	6.46	.710
	Average	6.24	.852
Perceived risk	PR1	1.98	.554
	PR2	1.94	.547
	PR3	1.98	.539
	PR4	2.04	.589
	PR5	2.11	.673
	Average	2.01	.580

(*Continued*)

Table 5. (*Continued*)

Construct	Item	Mean	Standard Deviation
Behavioural intention	BI1	6.70	.517
	BI2	6.71	.514
	BI3	6.70	.518
	Average	6.70	.516

adopting the M-Internet. This research included PR and INN to establish a suggested gap in UTAUT2 that disregarded the empirical findings about the significance of these two variables. Indeed, this is compatible with Venkatesh et al. [29] who suggest investigating UTAUT2 applicability on the M-Internet in different countries.

7 Conclusion

This study attempts to pinpoint the various factors that are considered to be effective on the behavioural intention of potential users when it comes to using the M-Internet in Saudi Arabia. Nine variables were selected, i.e. PE, EE, SI, FC, HM, PV, PR, INN and BI. Through running a descriptive analysis for every variable after giving the questionnaire to the participants, the findings of the analysis indicated that the related factors are effective when it comes to the behavioural intention of the participants.

7.1 Limitations and Future Research Directions

First of all, it will not be possible to generalise the findings to be representative of the overall population in Saudi Arabia because of the employment of a descriptive analysis rather than an inferential one which depends on using a Structural Equation Modelling (SEM). As this study considers investigating the behaviour intention instead of the actual use of the M-Internet, it is not going to provide a complete view about the actual usage of this service in Saudi Arabia. Moreover, conducting a longitudinal research is going to provide a better understanding regarding M-Internet in addition to clarifying the effects of the variables under study whose effects are stable and lasting over time. Consequently, any future study has to take into account these issues.

References

1. Agarwal, R., Prasad, J.: A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf. Syst. Res. **9**(2), 204–215 (1998)
2. Aldás-Manzano, J., Ruiz-Mafé, C., Sanz-Blas, S.: Exploring individual personality factors as drivers of M-shopping acceptance. Ind. Manage. Data Syst. **109**(6), 739–757 (2009)
3. Alkhunaizan, A., Love, S.: What drives mobile commerce? An empirical evaluation of the revised UTAUT model. Int. J. Manage. Mark. Acad. **2**(1), 82–99 (2012)

4. Alwahaishi, S., Snášel, V.: Consumers' acceptance and use of information and communications technology: A UTAUT and flow-based theoretical model. J. Technol. Manage. Innov. **8**(2), 61–73 (2013)
5. Alwahaishi, S., Snášel, V.: Factors influencing the consumers' adoption of mobile Internet. In: The Third International Conference on Digital Information and Communication Technology and its Applications DICTAP 2013, pp. 31–39. The Society of Digital Information and Wireless Communication (2013b)
6. Analysys Mason.: Final Report to the GSMA: the Socio-Economic Benefit of Allocating Harmonised Mobile Broadband Spectrum in the Kingdom of Saudi Arabia (2012). http://www.gsma.com/spectrum/wp-content/uploads/2012/05/GSMA_report_on_KSA_DD2_6GHz_2012-04-30.pdf. Accessed: 01 February 2015
7. Baabdullah, A.M., Dwivedi, Y.K., Williams, M.D.: IS/IT adoption research in the saudi arabian context: analysing past and outlining the future research directions. In: European, Mediterranean, and Middle Eastern Conference on Information System (EMCIS 2013), Windsor, United Kingdom, 17–18 October 2013 (2013)
8. Baabdullah, A. M., Dwivedi, Y. K., Williams, M. D.: Adopting an extended UTAUT2 to predict consumer adoption of M-technologies in Saudi Arabia. In: UK Academy for Information Systems Conference Proceedings 2014. Paper 5 (2014)
9. Brown, S.A., Venkatesh, V.: A model of adoption of technology in the household: A baseline model test and extension incorporating household life cycle. Manage. Inf. Syst. Q. **29**(3), 4 (2005)
10. Chae, M., Kim, J.: What's so different about the mobile Internet? Commun. ACM **46**(12), 240–247 (2003)
11. Chen, S.C., Li, S.H.: Consumer adoption of e-service: Integrating technology readiness with the theory of planned behavior. African J. Bus. Manage. **4**(16), 3556–3563 (2010)
12. Cheong, J.H., Park, M.C.: Mobile Internet acceptance in Korea. Internet Res. **15**(2), 125–140 (2005)
13. Dancey, C.P., Reidy, J.: Statistics Without Maths for Psychology. Pearson Education, Harlow (2007)
14. Dwivedi, Y.K., Wastell, D., Laumer, S., Henriksen, H.Z., Myers, M.D., Bunker, D., Elbanna, A., Ravishankar, M.N., Srivastava, S.C.: Research on information systems failures and successes: status update and future directions. Inf. Syst. Frontiers **17**(1), 143–157 (2015)
15. Dwivedi, Y.K., Choudrie, J., Brinkman, W.P.: Development of a survey instrument to examine consumer adoption of broadband. Ind. Manage. Data Syst. **106**(5), 700–718 (2006)
16. Ethos Interactive.: Internet and mobile users in Saudi Arabia KSA (2012). http://blog.ethosinteract.com/2012/02/06/Internet-user-in-saudi-arabia-ethos-interactive. Accessed: 03 February 2015
17. Featherman, M.S., Pavlou, P.A.: Predicting e-services adoption: a perceived risk facets perspective. Int. J. Hum Comput Stud. **59**(4), 451–474 (2003)
18. Fogelgren-Pedersen, A.: The mobile Internet: The pioneering users' adoption decisions. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, HICSS 2005, p. 84. IEEE (2005)
19. Hsu, H.H., Lu, H.P., Hsu, C.L.: Adoption of the mobile Internet: An empirical study of multimedia message service (MMS). Omega **35**(6), 715–726 (2007)
20. Huang, J., Liu, D.: Factors influencing continuance of mobile virtual community: Empirical evidence from China and Korea. In: 2011 8th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–6. IEEE (2011)
21. Jiang, P.: Adopting mobile Internet: A demographic and usage perspective. Int. J. Electron. Bus. **6**(3), 232–260 (2008)

22. Karaiskos, D.C., Kourouthanassis, P., Lantzouni, P., Giaglis, G.M., Georgiadis, C.K.: Understanding the adoption of mobile data services: Differences among mobile portal and mobile Internet users. In: 8th International Conference on Mobile Business, ICMB 2009, pp. 12–17. IEEE (2009)
23. Mallat, N., Rossi, M., Tuunainen, V.K., Oorni, A.: The impact of use situation and mobility on the acceptance of mobile ticketing services. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences, HICSS 2006, vol. 2, p. 42. IEEE (2006)
24. Shin, D.H.: The evaluation of user experience of the virtual world in relation to extrinsic and intrinsic motivation. Int. J. Hum.-Comput. Interact. **25**(6), 530–553 (2009)
25. Slade, E., Dwivedi, Y.K., Piercy, N.L., Williams, M.D.: Modeling consumers' adoption intentions of remote mobile payments in the UK: Extending UTAUT with innovativeness, risk and trust. Psychol. Mark. **32**(8), 860–873 (2015)
26. STC Group.: ICT in Saudi Arabia: A socio-economic impact review (2011). http://www.enlightenmenteconomics.com/Reports/assets/ICTinSaudi%20Arabia.pdf. Accessed: 05 February 2015
27. Thong, J.Y., Hong, S.J., Tam, K.Y.: The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. Int. J. Hum Comput Stud. **64**(9), 799–810 (2006)
28. Thong, J.Y., Venkatesh, V., Xu, X., Hong, S.J., Tam, K.Y.: Consumer acceptance of personal information and communication technology services. IEEE Trans. Eng. Manage. **58** (4), 613–625 (2011)
29. Venkatesh, V., Thong, J., Xu, X.: Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Q. **36**(1), 157–178 (2012)
30. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS Q. **27**, 425–478 (2003)
31. Williams, M.D., Rana, N.P., Dwivedi, Y.K.: The unified theory of acceptance and use of technology (UTAUT): A literature review. J. Enterp. Inf. Manage. **28**(3), 443–488 (2015)
32. Wiratmadja, I.I., Govindaraju, R., Athari, N.: The development of mobile Internet technology acceptance model. In: 2012 IEEE International Conference on Management of Innovation and Technology (ICMIT), pp. 384–388. IEEE (2012)
33. Wu, J.H., Wang, S.C.: What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. Inf. Manag. **42**(5), 719–729 (2005)
34. Zhou, T.: Examining location-based services usage from the perspectives of Unified Theory of Acceptance and Use of Technology and privacy risk. J. Electron. Commer. Res. **13**(2), 135–144 (2012)

Conceptualising and Exploring User Activities in Social Media

Marcel Rosenberger[1(✉)], Tobias Lehmkuhl[2], and Reinhard Jung[1]

[1] Institute of Information Management, University of St. Gallen, St. Gallen, Switzerland
marcel.rosenberger@student.unisg.ch, reinhard.jung@unisg.ch
[2] St. Gallen, Switzerland
tobias.lehmkuhl@swisscom.com

Abstract. A growing number of companies are recognising the benefits of using social media in customer relationship management. At the same time, the consumers' expectations are rising: short response times, individual communication, real interaction with humans, and participation. It is a challenge to observe the many different user activities on many different social media sites. The aim is to reduce the complexity of integrating multiple social media sites with enterprise systems. Therefore, a conceptualisation of user activities in social media is presented. A user activity is a cross-over of an action invoked on an object and a user who acts in a certain context. The 40 user activity types are compared with actual features of ten social media sites. We find out that a substantial share of them can be integrated technically using the social media site's Application Programming Interfaces (APIs).

Keywords: Social media · User activities · Conceptualisation · Integration

1 Introduction

Social media have become a noticeable part of society. This development attracts attention of companies that aim to take advantage of the opportunities, such as improvement of the reputation and marketing efficiency, support cost reduction, and product innovation from co-creation [1–5]. At the same time, consumers benefit from participating companies, e.g. through relationship advantages, interaction and exchange, and influence on business processes. Examples are discounts, special promotions, and the acceleration of the fulfilment of support requests. The many active users perform various actions in social media and create a lot of data therewith. This information overload is a challenge for companies, because the increasing number of content, user profiles, and connections cannot be timely assessed manually [1].

Information systems (IS) are needed to manage the social media initiatives, providing functions to publish, observe and analyse social media data and integrate it with company data [6]. A preliminary step is to identify business-relevant user activities and to process the related data. Examples of user activities are joining groups, placing like- or dislike flags, adding others to the friend list, reading specific texts, watching videos, and changing

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 107–118, 2015.
DOI: 10.1007/978-3-319-25013-7_9

profile information. These user activities comprise business opportunities in the form of leads (i.e. potential customers), enriched customer profile information, and a better understanding of interests and markets. Fliess & Nesper [4] state that "activities of customers can be considered as an economic resource". Similarly, Holts [7] highlights that user activities in social media create value and stimulate companies' revenues.

The more social media sites are considered by the company, the higher is the media penetration. Consequently, there is a need to integrate multiple social media sites mutually. However, the social media sites are diverse and facilitate different user activities. Posts, tweets, pins, profiles, groups, and pages, which are posted, tweeted, pinned, modified, added, or viewed, are only a small proportion. There are no common social media data structures, on which the integrations could be built.

Research on user activities in social media is contemporary and there are a number of existing conceptualisations [17–19, 22]. These are valuable to understand the user's motivation of being active and show some features of social media. However, the proposals are inappropriate to guide the implementation of integration software between multiple social media and an enterprise system. For this purpose, the existing conceptualisations are too abstract or they are exemplary and not exhaustive. Especially the related data of the user activities and the technical accessibility had not been researched yet. Our proposed conceptualisation and exploration of user activities closes this gap.

The topic is motivated from practitioners. The research is part of a joint social Customer Relationship Management (CRM) program with scientists and practitioners from companies of the insurance industry. The status quo of the companies shows that social media tools are isolated and not technically integrated into existing CRM systems. Social media monitoring tools are used to capture developments and to extract aggregated metrics, such as number of posts, likes, and age distribution on single sites. Relevant posts can be identified automatically based on tags, keywords, and rules. However, software solutions that recognise business-relevant user activities in multiple social media and invoke adequate business processes automatically are not yet implemented. Profiting business areas are customer service, sales, marketing, human resources, and research and development [3]. A conceptualisation of user activities is useful for designing general integration solutions. The audience are practitioners and researchers of social media and information systems.

Chapter 2 gives the conceptual background. Chapter 3 connects to existing knowledge and related work. The research methodology is presented in Chap. 4. Then, the user activity types in social media are described (Chap. 5) and compared with actual features of ten social media sites (Chap. 6). The final chapter states implications, limitations, and guides further research.

2 Related Work

Web 2.0 is an economic, social, and technology concept of the Internet, which enables users to create content and build a network with other users [8]. The results from user participation, e.g. posts, friend lists, and profiles, are accessible by other parties of the community. As stated in the definition by Kaplan & Haenlein [9], "social media is a group of Internet-based applications that build on the ideological and technological

foundations of web 2.0". We use the term to refer to the sites/platforms that are built on the web 2.0 concept (e.g. Facebook, Google +, and Twitter). The terms "social media site" or "social media sites" are only used when an emphasis on singular or plural is necessary. A basic principle of social media is to connect to others and share information [10]. Social media and enterprise systems are heterogeneous systems, which can be connected through system integration. According to Hasselbring [11], heterogeneity leads to complexity, which is an issue for the integration task.

Küpper et al. [12] show results from a market study of 40 vendor solutions for social media tools. The findings indicate that most tools provide features to capture and analyse aggregated social media data. The capturing and analysis of individual data (i.e. single posts, user profiles, etc.) as well as the integration into enterprise systems is sparse. Similarly, other authors state that the integration of social media with enterprise systems is still insufficient [13, 14]. In particular, Trainor et al. [15] identify a lack of interaction between CRM systems and social media technology. For example, customer data and user data in social media are not interrelated and business-processes are not triggered from incidents in social media automatically.

Atig et al. [16] conceive user activity as the time when the user is active in social media. The authors classify users based on activity profiles and thereby do not differentiate between what the users are actually doing when they are active. Heinonen [17] conceptualises consumers' social media activities based on two dimensions: consumer motivation and consumer input. The consumers' motivation to use social media falls into one of three categories: information processing, entertainment activities, and social connection. The consumer input has three main types, which are consumption, participation, and contribution. The author's framework allows classifying users' activities. For example, "creating and managing a social network" is motivated from the need for social connection and requires creating a profile and linking to friends (productive consumer input). The proposed framework is abstract and does not allow deriving the related data of the activities. Pankong et al.'s [18] ontology for social activities is more concrete. In principle, the ontology is an entity-relationship-model, which shows entities (e.g. users, posts, likes, and topics) and its relationships (e.g. "is a", "has a", and "related to"). Some entities, however, are ambiguous (e.g. reply, retweet, and comment). Besides, the viewing of content is not included in the ontology. The model facilitates a snapshot-view of the social media graph. The circumstances in which the users create the content is not incorporated. This is justifiable considering that the authors focus on existing explicit and implicit relationships of users, similarly to Yang et al. [19]. However, the location and time of an activity are also expedient to determine the business-relevance [20]. Hotho & Chin [21] analyse the circumstances of user activities. Available sensors of a smartphone are used to conceive the current situation of the user (e.g. installed applications, busy status, missed calls count, position from Graphical Position System (GPS) sensor, remaining battery power, and ringtone volume). Richthammer et al. [22] identify 11 online social network (OSN) activities. Examples are "User posts Item/Comment", "User sends messages to Contact/Page", "User is linked to Item/Comment", and "Contact/Page views User's Profile". However, these are only "fundamental user activities on OSNs" and are not complete. For example, the sharing, deletion and modification of content is not considered.

3 Methods

The literature review follows vom Brocke et al.'s [23] methodology, which comprises three process steps, being (1) definition of review scope, (2) conceptualisation of topic, and (3) literature search. The authors highlight that not only results should be presented, but, to allow replicability, also details about the approach. The scope (1) of the literature review is characterised by six aspects borrowing from Cooper [24] (Table 1).

Table 1. Taxonomy of the conducted literature review (borrowing from [24])

Characteristic	**Categories**			
(a) focus	research outcomes	research methods	theories	applications
(b) goal	integration	criticism	central issues	
(c) organisation	historical	conceptual	methodological	
(d) perspective	neutral representation	espousal of position		
(e) audience	specialised scholars	general scholars	practitioners	general public
(f) coverage	exhaustive	exhaustive and selective	representative	central/pivotal

The focus (a) is on existing research results concerning user activities in social media. The goal (b) is to connect to existing knowledge on a conceptual level (c). The perspective (d) can be characterised as neutral representation, because the position is unbiased. Practitioners and researchers of social media are the target audience (e). The results are representative (f) for the IS community, because prominent data sources have been queried.

The conceptualisation of the topic (2) includes a "working definition of [the] key variable(s)" [25]. A keyword search (3) in the databases of AISel, EBSCO, Emerald, IEEE, JSTOR, ProQuest, and Web of Science in the title (TI), topic (TO), abstract (AB), keyword (KW), and full text (TX) fields was applied using the search string: *"social media" AND ("user actions" OR "user activities")*. The initial list of publications has been filtered by reading the titles and abstracts. Relevant papers were analysed based on the full texts. Table 2 shows the numerical results of the keyword searches.

The development of the user activity types comprised a study of features of large, popular social media. The sample of sites for analysis has been selected on the following criteria: (1) large number of active users per month (>100 m.); (2) English localisation of the platform; (3) availability of a public API; and (4) permission for commercial use. The initial list of contemplable sites has been compiled of studies and rankings of social media [26, 27]. The listed sites have been evaluated against the aforementioned criteria, based on information from press releases, technical notes, terms of use, and responses from enquiries to the providers. Possible user activities

Table 2. Numerical results of the keyword searches

Data source	Search fields	Publications	
		Total	Relevant
AISeL	TI, AB	11	2
EBSCO	TI, AB, KW, TX	22	5
Emerald	TI, AB, KW	30	1
IEEE	TI, AB, KW	4	1
JSTOR	TI, AB, KW, TX	9	1
ProQuest	TI, AB, KW	4	1
Web of Science	TI, TO	4	1
Total[a]			**10**

[a] The total number is not equal to the column sum, because duplicates have been counted only once.

have been gathered by analysis of the features and functions. They have been grouped according to the philosophical idea of family resemblance and following an abstraction-based modelling approach [28, 29].

4 User Activity Types in Social Media

The user activity types shown in Fig. 1. represent the actions that users perform in social media. Activities take place in a context, in which the user is situated, defined by time, location, social media site, device, and application. The combination of an object type and an action is termed a user activity type in social media. The complex graph structure of social media is broken down into an activity log, which contains entries of the form: user u invoked action a on object o (on site s with device d in application p from location l at time t).

Five actions can be applied to eight object types. All user-generated content results from the Create-, Update-, or Share-action. The content is displayed on the screens of the users' devices via the View-action. The Delete-action removes content. The variety of features across different social media that facilitate the creation, modification and viewing of content is reducible to 8×5 user activity types.

4.1 Social Media Object Types

The idea of family resemblance is adduced to group similar objects. The most prototypical objects constitute an object type. An object type subsumes all objects, which have most functions and structure in common with that object type, and have least commonalities with other object types [28].

Kietzmann et al. [30] present a framework of functional building blocks of social media, which are identity, conversations, sharing, presence, relationships, reputation, and groups. The seven blocks are facets of user experience in social media and give an orientation to gather object types and functions. Table 3 identifies social media object types by analysing its structure and functions [31].

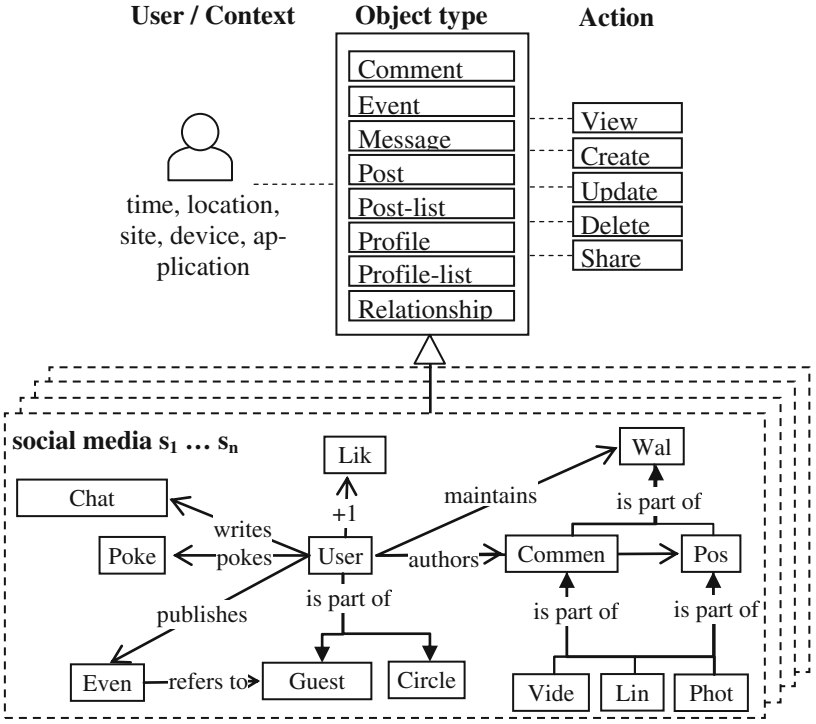


Fig. 1. User activity types in social media: a combination of context, object type and action

An object type is enabling a function, if it supports the intention behind the concept of the building block. It is partly enabling, if the intended user experience of the functional building block is a side-effect. An object type is not contributing to a functional building block, if it does not enable the function.

The eight object types may have variants that share similar concepts, but have a different terminology or are used to distinguish the same concept that is used in different contexts on the same site. An example is a Comment, which contains multimedia content and has a parent, which may be a Post, Post-list, Comment or another object type. Some sites use a Comment object type to represent answers, like Tumblr, or a job application, like Xing.

The same object type is also used in different variants on the same social media site, such as on Facebook, where both comments and reviews exist. A Comment primarily contributes to the functional building blocks Conversations and Sharing. An Event object type defines a happening, which has relatedness to time. It can be a birthday party, a music festival, a meeting, and so on. Events facilitate to meet people (Presence), build communities (Groups), and relate to each other (Relationships). A Message is multimedia content that is addressed to a specified set of receivers. Posts subsume a main content entry found on all social media sites. They may be termed tweet, job, or pin, and engage interaction by allowing adding Comments and Relationships.

Table 3. Technical identity of social media objects

Object types	Structure / data	Functions						
		Identity	Conversations	Sharing	Presence	Relationships	Reputation	Groups
Comment	- contains content (e.g. text, video, audio and image) - refers to another object	○	●	●	◐	○	◐	○
Event	- contains descriptive information about a happening - has a relatedness to time	○	○	○	●	●	◐	●
Message	- contains content (e.g. text, video, audio and image) - has a sender and recipient (list)	○	●	●	◐	○	○	○
Post	- contains content (e.g. text, video, audio and image)	○	●	●	◐	○	◐	○
Post-list	- is a collection of posts	○	●	●	◐	◐	○	◐
Profile	- contains descriptive information about an actor	●	○	◐	●	◐	●	○
Profile-list	- is a collection of profiles	◐	◐	○	◐	●	●	●
Relationship	- connects two objects	○	◐	○	◐	●	●	◐

Legend: Object type is … ○ not enabling the function, ◐ partly enabling the function, ● enabling the function

A Post-list is a collection of Posts. A Profile is a representation of an entity of the real life, such as a person, company, or a community. A Profile-list is a collection of Profiles with possible variants, such as circle, contact list, and friend list. A Relationship connects two other objects. An example is a bookmark, which can be described as a Relationship between a Profile and a Post. The poke feature in Facebook can be treated as a Relationship between two Profiles.

4.2 Actions on Social Media Objects

Table 4 lists actions, which can be invoked on social media objects referring to Hypertext Transfer Protocol (HTTP) methods. HTTP is the underlying, technological protocol of social media sites [32]. Five basic actions on social media objects can be

Table 4. Actions on social media objects

Action	Description	HTTP methods
View	View is triggered when content of an object is loaded and displayed on the user's screen (e.g.: a video is played)	GET
Create	The Create-action occurs when something new is added opposed to the Update-action when a change to an existing object is done by a user	POST
Update	The Update-action results in a modified, existing social media object	PUT/MOVE
Delete	When an object is removed on social media an event with action Delete is raised	DELETE
Share	The Share-action occurs when existing content, usually originated from another user, is put into a different context or is exposed to additional users on the same platform. It is a copy of an already existing entity	COPY

identified. Sharing is something particularly found in social media [30]. The citing of a text phrase or the re-tweet of content on Twitter is an example of the Share-action.

5 Empirical Exploration

Table 5 shows the results of the empirical exploration of the user activities in ten social media.

The APIs of large social media define access options to functions and data using web-services. They include formats and provide methods to publish posts, resolve connections between users, and retrieve comments, for example. Dark underlined numbers signify that the user activity type can be monitored in the specified social media using the provided API. Dark numbers that are not underlined mean that the type exists on the site, but the APIs of the site do not provide access to monitor it. For example, in Facebook a user can view a post. However, this activity cannot be monitored using the public API of Facebook in the recent version of the Graph API V2.1 [33]. On the other hand, it cannot be ruled out that access is included in upcoming versions. Furthermore, using the APIs is not the only access approach to social media data. Instead of using the API, the View-action of own and shared posts can be recognised by linking a Facebook post with external content from a corporate website, where the company can evaluate page requests (by observing the HTTP/1.1 GET-method).

Six user activity types are theoretical constructs, which do not occur in the analysed social media. These are Message/Update, Message/Delete, Message/Share, Profile-list/Share, Relationship/Update, and Relationship/Share. Firstly, a Message is private, because it cannot be shared. Secondly, a Message, once sent, cannot be fetched back, removed, or edited. A Profile-list cannot be shared by others. Access privileges of the Profile-lists are maintained by the owners only. A Relationship does either exist or does not exist. It cannot be modified; but it can be deleted.

Table 5. Empirical exploration of user activities in social media

Object type \ Action	View	Create	Update	Delete	Share
Comment Answer, Recommendation, Job Application, Review	❶❷❸❹❺ ❻❼❽❾❿	❶❷❸❹❺ ❻❼❽❾❿	❶②❸④❺ ⑥⑦⑧❾❿	❶②❸❹❺ ❻❼❽❾❿	❶②③④⑤ ⑥❼❽❾⑩
Event Meeting, Happening	❶②❸④❺ ⑥⑦⑧❾⑩	❶②❸④❺ ⑥⑦⑧❾⑩	❶②❸④❺ ⑥⑦⑧❾⑩	❶②❸④❺ ⑥⑦⑧❾⑩	❶②❸④⑤ ⑥⑦⑧❾⑩
Message Chat, Fanpost, Gift	❶❷❸❹❺ ⑥❼❽❾⑩	❶❷❸❹❺ ⑥❼❽❾⑩	①②③④⑤ ⑥⑦⑧⑨⑩	①②③④⑤ ⑥⑦⑧⑨⑩	①②③④⑤ ⑥⑦⑧⑨⑩
Post Job, Life event, Pin, Project, Question, Status, Tweet	❶❷❸❹❺ ❻❼❽❾⑩	❶❷❸❹❺ ❻❼❽❾⑩	❶②❸④❺ ❻❼⑧❾⑩	❶❷❸❹❺ ❻❼❽❾⑩	❶②❸❹⑤ ❻❼❽❾⑩
Post-list Blog, Board, Page, Photo album, Wall	❶❷❸❹❺ ❻❼⑧❾⑩	❶❷❸❹❺ ❻❼⑧❾⑩	❶❷❸❹❺ ❻❼⑧❾⑩	❶❷❸❹❺ ❻❼⑧❾⑩	❶②❸④⑤ ⑥⑦⑧❾⑩
Profile Community, Company, User	❶❷❸❹❺ ❻❼❽❾⑩	❶❷❸❹❺ ❻❼❽❾⑩	❶❷❸❹❺ ❻❼❽❾⑩	❶❷❸❹❺ ❻❼❽❾⑩	①②③④⑤ ⑥⑦⑧❾⑩
Profile-list Circle, Contact list, Friend list, Group, Guest list, Partner list	❶❷❸❹❺ ⑥⑦❽❾⑩	❶❷❸❹❺ ⑥⑦❽⑨⑩	❶②❸❹❺ ⑥⑦❽❾⑩	❶②❸❹❺ ⑥⑦❽⑨⑩	①②③④⑤ ⑥⑦⑧⑨⑩
Relationship Bookmark, Favourite, Follow, Invitation, Join, Like, Poke, Rating	❶❷❸❹❺ ❻❼❽❾⑩	❶❷❸❹❺ ❻❼❽❾⑩	①②③④⑤ ⑥⑦⑧⑨⑩	❶❷❸❹❺ ❻❼❽❾⑩	①②③④⑤ ⑥⑦⑧⑨⑩

Legend: 1-Facebook, 2-Flickr, 3-Google+, 4-LinkedIn, 5-Meetup, 6-Pinterest, 7-Tumblr, 8-Twitter, 9-Xing, 10-YouTube
Event type is … ○ not existent on the social media site, ● existent on the social media site, ● existent on the social media site, and can be accessed using the API

The majority (70 %) of user activity types that exist on a social media site, can also be monitored using the API and thus can be integrated with enterprise systems using a public, recommended access approach. The View-actions are usually not provided; only Google + has custom activities, which can be triggered by developers in case an

entity is read. On most social media Post/Create is observable by subscription to Post-lists. The Update-, and Delete-actions could be identified by periodic polling, whereby known objects are checked regularly to notice if they are still existent or modified. Comment, Post, Profile, and Relationship exist in all analysed social media. Thus, these are essential object types. Facebook (83 %), Google + (80 %), and Xing (80 %) feature the most complete set of user activity types. The APIs of Google +, LinkedIn, and YouTube provide the most complete set of access options, covering 95 %, 80 %, and 79 % of applicable user activities of each site.

6 Discussion and Conclusions

The user activity types define user activities in social media. They specify what users do in social media when they create or consume content. Hence, the user activity types advance from existing definitions of user activities that conceive user activity as the time when the user is active in social media [16]. A user activity type is a crossover of a social media object type and an action and takes place in a specific user's context. The object types reveal the underlying structure and data, which large social media sites share. The actions are operations that users perform with an object type. The user-context describes the situation in which the user resides while invoking an action on an object.

The results are useful to design and develop integration software that facilitates to process user activities of multiple different social media sites. Middleware-based solutions require similar structuring of information. The presented user activity types support that purpose, because they allow to consolidate the different user activities of different, large social media sites. There are technical restrictions limiting the feasibility to capture "everything", because some user activity types cannot be captured using the APIs. Moreover, as also highlighted by other authors, users' permission and privacy need to be considered [34]. It must be a major concern of all business-oriented social media initiatives, because of the risk to destroy relationships to customers in case of an accident. An example is unintended data exposure to unauthorised parties. As a result, not every user activity that can be monitored technically should also be tracked.

The user activity types originate from the abstraction of individual features collected from a study of ten social media sites. They have an empirical basis and rely on publicly available data. The issue, caused by the underlying induction of the abstraction, is that the user activity types are only certainly valid for the analysed social media, and are not necessarily generalisable to all available sites.

Further research is encouraged to concretise the user activity types in terms of a canonical data schema, which defines data types and attributes. Based on the detailed level, (business-specific) rules can be proposed for filtering user activities. Monitoring of user activities in social media leads to a reactive system [35]. A fully integrated IS, however, should comprise functions to interact, requiring both directions of a communication. This is not contrary to the research results, but is a possible extension.

References

1. Smith, T.: Conference notes – the social media revolution. Int. J. Mark. Res. **51**, 559 (2009)
2. Baird, C.H., Parasnis, G.: From social media to social customer relationship management. Strateg. Leadersh. **39**, 30–37 (2011)
3. Cappuccio, S., Kulkarni, S., Sohail, M., Haider, M., Wang, X.: Social CRM for SMEs: current tools and strategy. In: Khachidze, V., Wang, T., Siddiqui, S., Liu, V., Cappuccio, S., Lim, A. (eds.) iCETS 2012. CCIS, vol. 332, pp. 422–435. Springer, Heidelberg (2012)
4. Fliess, S., Nesper, J.: Understanding patterns of customer engagement – How companies can gain a surplus from a social phenomenon. J. Mark. Dev. Compet. **6**, 81–93 (2012)
5. Jahn, B., Kunz, W.: How to transform consumers into fans of your brand. J. Serv. Manag. **23**, 344–361 (2012)
6. Acker, O., Gröne, F., Akkad, F., Pötscher, F., Yazbek, R.: Social CRM: How companies can link into the social web of consumers. J. Direct Data Digit. Mark. Pract. **13**, 3–10 (2011)
7. Holts, K.: Towards a Taxonomy of Virtual Work. In: 31st International Labour Conference 2013, vol. 7, pp. 31–50. Rutgers Univ. New Brunswick, New York (2013)
8. Musser, J., O'Reilly, T.: Web 2.0 Principles and Best Practices. O'Reilly Media, Sebastopol (2006)
9. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media. Bus. Horiz. **53**, 59–68 (2010)
10. Ang, L.: Is SCRM really a good social media strategy? J. Database Mark. Cust. Strateg. Manag. **18**, 149–153 (2011)
11. Hasselbring, W.: Information system integration. Commun. ACM **43**, 32–38 (2000)
12. Küpper, T., Lehmkuhl, T., Jung, R., Wieneke, A.: Features for social CRM technology – An organizational perspective. In: AMCIS 2014 Proceedings, pp. 1–10 (2014)
13. Sarner, A., Thompson, E., Sussin, J., Drakos, N., Maoz, M., Davies, J., Mann, J.: Magic Quadrant for Social CRM, pp. 1–20 (2012)
14. Reinhold, O., Alt, R.: How companies are implementing social customer relationship management: insights from two case studies. In: Proceedings of the 26th Bled eConference, pp. 206–221 (2013)
15. Trainor, K.J., Andzulis, J., Andzulis, J.M., Rapp, A., Agnihotri, R.: Social media technology usage and customer relationship performance: a capabilities-based examination of social CRM. J. Bus. Res. **67**, 1201 (2013)
16. Atig, M.F., Cassel, S., Kaati, L., Shrestha, A.: Activity profiles in online social media. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), pp. 850–855 (2014)
17. Heinonen, K.: Consumer activity in social media: managerial approaches to consumers' social media behavior. J. Consum. Behav. **10**, 356–364 (2011)
18. Pankong, N., Prakancharoen, S., Buranarach, M.: A combined semantic social network analysis framework to integrate social media data. In: Proceedings of 2012 4th Internatioanl Conference Knowledge Smart Technology KST 2012, pp. 37–42 (2012)
19. Yang, C.C., Tang, X., Dai, Q., Yang, H., Jiang, L.: Identifying implicit and explicit relationships through user activities in social media. Int. J. Electron. Commer. **18**, 73–96 (2013)
20. Yu, Y., Tang, S., Zimmermann, R., Aizawa, K.: Empirical observation of user activities. In: Proceedings of the First International Workshop on Internet-Scale Multimedia Management - WISMM 2014, pp. 31–34 (2014)

21. Woerndl, W., Manhardt, A., Schulze, F., Prinz, V.: Logging user activities and sensor data on mobile devices. In: Atzmueller, M., Hotho, A., Strohmaier, M., Chin, A. (eds.) MUSE/MSM 2010. LNCS, vol. 6904, pp. 1–19. Springer, Heidelberg (2011)
22. Richthammer, C., Netter, M., Riesner, M., Sänger, J., Pernul, G.: Taxonomy of social network data types. EURASIP J. Inf. Secur. **2014**, 11 (2014)
23. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., Von Brocke, J., Reimer, K.: Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: 17th European Conference on Information Systems (2009)
24. Cooper, H.M.: Organizing knowledge syntheses: a taxonomy of literature reviews. Knowl. Soc. **1**, 104–126 (1988)
25. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **26**, xiii–xxiii (2002)
26. Singh, N., Lehnert, K., Bostick, K.: Global social media usage: insights into reaching consumers worldwide. Thunderbird Int. Bus. Rev. **54**, 683–700 (2012)
27. Statista: Leading social networks worldwide as of June 2014. ranked by number of active users. http://www.statista.com/statistics/272014/global-social-networks-ranked-by-num-ber-of-users/
28. Rosch, E., Mervis, C.B.: Family resemblances: studies in the internal structure of categories. Cogn. Psychol. **7**, 573–605 (1975)
29. Bussler, C.: Modeling Methodology. B2B Integration: Concepts and Architecture. Springer, Heidelberg (2003)
30. Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S.: Social media? Get serious! understanding the functional building blocks of social media. Bus. Horiz. **54**, 241–251 (2011)
31. Faulkner, P., Runde, J.: Technological objects, social positions, and the transformational model of social activity. MIS Q. **37**, 803–818 (2013)
32. Gourley, D., Totty, B., Sayer, M., Aggarwal, A., Reddy, S.: HTTP: The Definitive Guide. O'Reilly Media, Inc, California (2002)
33. Facebook: Graph API Reference. https://developers.facebook.com/docs/graph-api/reference/v2.1
34. Woodcock, N., Broomfield, N., Downer, G., McKee, S.: The evolving data architecture of social customer relationship management. J. Direct Data Digit. Mark. Pract. **12**, 249–266 (2011)
35. Sarkar, A., Waxman, R., Cohoon, J.P.: High-level system modeling. In: Bergé, J.-M., Levia, O., Rouillard, J. (eds.) High-Level System Modeling: Specification Languages, pp. 1–34. Springer, US (1995)

Understanding the Determinants of Privacy-ABC Technologies Adoption by Service Providers

Ahmad Sabouri(✉)

Deutsche Telekom Chair of Mobile Business and Multilateral Security,
Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany
ahmad.sabouri@m-chair.de

Abstract. As using online services penetrates deeper in our everyday life, lots of trust-sensitive transactions are carried out electronically. In this regard, a big challenge is to deal with proper user authentication and access control without threatening the users' privacy. However, commonly used strong authentication schemes fail to address important privacy requirements. In this paper, we focus on an emerging type of digital certificates, known as Privacy-preserving Attribute-based Credentials (Privacy-ABCs), which allow privacy and security go hand-in-hand. So far, there has been no systematic study on the potential factors that have influence on the adoption of Privacy-ABCs by service providers. Thus, we developed a conceptual model of the relevant factors based on well-established theories and our practical experience with trialing Privacy-ABCs, and evaluated the model through expert surveys.

Keywords: Privacy-preserving attribute-based credentials · Anonymous credentials · Technology adoption · Expert surveys

1 Introduction

Nowadays, usernames and passwords are the most commonly used authentication schemes. However, the hassle of managing different usernames and passwords grows as the number of electronic services increases. This, on the one hand, raises security risks because many users tend to reuse the same password for different services. On the other hand, it introduces privacy threats for cross-linking activities of the users in different domains as it is highly probable to be able to correlate different identifiers of the same person [35] because they typically prefer to choose the same or similar usernames for their various accounts.

An alternative solution to improve the security problem is to employ strong authentication techniques such as digital certificates. Nonetheless, the most commonly used strong authentication techniques do not follow the Privacy-by-Design [6] principle of *Data Minimization*. For instance, the use of X509 certificates causes "Over-Identification", as it mandates the users to reveal all the attested

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 119–132, 2015.
DOI: 10.1007/978-3-319-25013-7_10

attributes in the certificate so that the validity of the digital signature is preserved, even if only a subset of attributes is required for the authentication purpose. Using online federated authentication and authorization techniques such as OpenID, SAML, Facebook Connect, and OAuth could support the minimal disclosure principal and allow the users to provide the service providers with only the requested information rather than the whole user's profile stored at the Identity Service Provider (IdSP). However, all these protocols suffer from a so-called "Calling Home" problem, meaning that for every authentication transaction the user is required to contact the IdSP (e.g., Facebook, Gmail, OpenID Provider). This introduces privacy risks to both users and service providers.

The focus of this paper is on a promising type of digital certificates called Privacy-preserving Attribute-based Credentials (Privacy-ABCs) that provide a strong basis for secure yet privacy-enhanced access control systems. Privacy-ABCs offer a solution to cope with *Minimal Disclosure* of attributes as well as supporting *Partial Identities*. Privacy-ABC users can obtain credentials from their IdSPs and when authenticating to different service providers, they can produce *unlinkable Privacy-ABC tokens* containing only the required subset of information available in the credentials without involving the IdSP or any third party in the process. Therefore, they can help overcome the "Over-Identification" and "Calling Home" problems. The prominent instantiations of such Privacy-ABC technologies are Microsoft U-Prove[1] and IBM Idemix[2]. Both of these technologies are studied in depth by the EU-funded project ABC4Trust, where a common architecture for Privacy-ABCs was designed, implemented and verified in two real-life trials [27].

Privacy-ABCs are emerging technologies that are not yet properly adopted. There have been a handful of proposals on how to realize a Privacy-ABC system in the literature [4,5]. However, the diversity of their features and implementations hindered their practical use. As Privacy-ABCs are in the pre-adoption phase, our rigorous literature review on drivers and inhibitors of Privacy-ABCs using well-known databases such as JStore, MISQ, AISnet, ACM and IEEE ended up in a limited set. Borking investigated the adoption of Privacy Enhancing Technologies (PETs) in general [3]. Nevertheless, PETs can be very different in their characteristics and their adoption schemes. For instance, Tor[3] is also an example of PETs that can be employed directly by the end users, while in order to have Privacy-ABC technologies operational, at least three entities have to adopt or accept the technology: (1) Credential issuers, which are typically organizations such as governments, banks, and telco operators who have authentic source of data about the users, (2) Service providers, which perform access control to their resources relying on the credential attested by the issuers, (3) Users, who consider using such kind of credentials. Therefore, Privacy-ABCs have special characteristics and effects that make them deserve a separate study. Therefore, this paper focuses on the adoption factors influencing **service providers** and

[1] http://microsoft.com/uprove.
[2] http://idemix.wordpress.com/.
[3] https://www.torproject.org/.

launches the first systematic work based on well-established theories to investigate the (future) adoption of Privacy-ABCs. Understanding the determinants of adoption by service providers is very important in the sense that identifying these factors can facilitate building guidelines for the supporting bodies to pave the road for the further adoption of Privacy-ABCs. Therefore, we developed a conceptual model based on the existing innovation adoption theories and evaluated the factors through expert surveys.

The rest of this paper is organized as follows. In Sect. 2, we introduce the features and concepts of Privacy-ABCs in more details and also deliver an overview of the theories in the literature explaining innovation adoption. Later, we present our conceptual model of the determinants in Sect. 3. Then, in Sect. 4, we present our empirical evaluation of the factors, and later in Sect. 5 discuss our findings and their implications. In the end, we conclude the paper in Sect. 6.

2 Theoretical Background

2.1 How Privacy-ABCs Work

A *Credential* is defined to be "a certified container of attributes issued by a credential Issuer to a User" [1]. An *Issuer* vouches for the correctness of the attribute values for a *User* when issuing a credential for her. In an example scenario, Alice as a *User*, contacts the Bundesdruckerei (the German authority responsible for issuing electronic IDs) and after a proper proof of her identity (e.g. showing her old paper-based ID), she receives a digital identity credential containing her first name, surname and birth-date. In the next step, she can seek to access an online Discussion Forum. The service provider provides Alice with the access policy that requires her to deliver an authentic proof of her first name. Using Privacy-ABCs features, Alice has the possibility to derive a minimal authentication token from her identity credential that contains only the first name. As a result, her privacy is preserved by not disclosing unnecessary information (i.e. surname and birth-date). Note that the commonly used digital certificates do not offer such capability as any change in those certificates invalidates the issuers' signature. Another example where Alice could use her Privacy-ABC might be with an online movie rental website, which requires age verification. Alice is able to provide such a proof without actually disclosing her exact birth-date. The proof is done based on complex cryptographic concepts that can show her birth-date attribute in her credential is before a certain date.

2.2 Innovation Adoption

A prominent approach to investigate adoption of new technologies is covered by the Diffusion of Innovation theory (DOI), presented by Rogers [26]. DOI theory sees innovations as being communicated through certain channels over time and within a particular social system. The approach focuses on the way in which a new technological invention migrates from creation to use. Rogers identified

five important attributes of innovations that might influence the decision for their adoption or rejection. The five characteristics of innovations are relative advantage, compatibility, complexity, trialability, and observability, which are valid for both individual and organizational adoption of technology.

Technology-Organization-Environment (TOE) framework was presented by Tornatzky [33] to study the adoption of technological innovations. The framework considers a threefold context for adoption and implementation of technological innovations: technological context, organizational context, and environmental context. The technological context relates to the technologies relevant to the firm such as the current internal practices and equipment, as well as the set of relevant technologies external to the firm. The organizational context describes the characteristics of an organization including firm size, degree of centralization, formalization, complexity of its managerial structure, the quality of its human resources, and the amount of slack resources available internally. Comparing to Rogers' model, TOE includes a new and important component, environmental context. The environment context is the arena in which a firm conducts its business such government and the competitors.

Iacovou et al. [15] presented a model to investigate the interorganizational systems (IOSs) characteristics that influence firms to adopt IT innovations in the context of Electronic Data Interchange (EDI). In this model, Perceived Benefits is a different factor from the TOE framework, whereas organizational readiness is a combination of the technology and organization context of the TOE framework. Nevertheless, Iacovou et al. included and highlighted external pressure as an important factor.

We have identified the Institutional Theory also to be relevant for our research. Institutional factors including schemas, rules, norms, and routines are crucial in shaping organizational structure and organizational decisions [28]. According to the institutional theory, organizational decisions are not driven purely by rational goals of efficiency, but also by social and cultural factors and concerns for legitimacy. It is posited by DiMaggio and Powell [9] that Coercive isomorphism, known as the pressures from other organizations, Mimetic isomorphism, known as the imitation of structures adopted by others in response to pressures, and Normative isomorphism, known as conformity to normative standards established by external institutions, potentially have influence on the behaviour of an organization.

3 Conceptual Model for Adoption of Privacy-ABCs

Based on the theories explained in the previous section, we constructed a combined conceptual model of the relevant factors that are potentially applicable to Privacy-ABCs adoption. We also propose some factors that are new and specific to the domain of Privacy and characteristics of Privacy-ABCs. The conceptual model presented in Fig. 1 incorporates thirteen factors categorized in five groups.

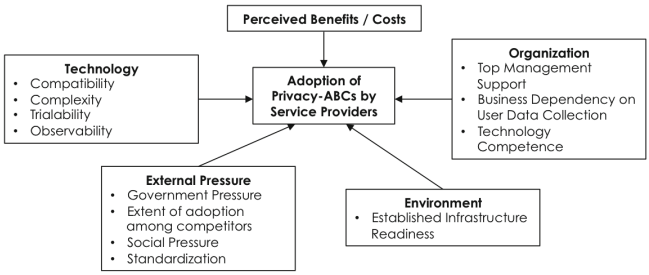


Fig. 1. Conceptual model for factors influencing adoption of Privacy-ABCs by service providers

Technology

Privacy-ABCs are a kind of new technologies for privacy-respecting access control and their characteristics may have a strong influence on the decision of the potential adopters.

Compatibility: Refers to the degree to which an innovation is perceived as consistent with the existing values, needs, and past experiences of the potential adopters [26]. Various published studies examined the role of compatibility, and considered it to be an essential determinant of IT innovation adoption [23,31,32,36] and several of them (e.g. [32,36]) found it as a significant driver. Regarding Privacy-ABCs, it is probable that a higher compatibility of their specifications with the existing Identity and Access Management (IAM) protocols and standards that are commonly used increases the likelihood of a positive decision to adopt them, as any change in the IAM processes may affect a wide range of the subsystems.

Complexity: Refers the degree to which an innovation is perceived as relatively difficult to understand and use [26]. There have been several works considering the role of complexity in innovation adoption [7,16,32]. Privacy-ABCs are based on difficult cryptographic concepts, which are not easy for people beyond the inventors to understand. On the one hand, following the claim by Borking [3], understanding of Privacy-ABCs for the purpose of adoption requires technical and legal knowledge. On the other hand, Wästlund et al. [34] claim that the users have difficulties in using these novel technologies. Therefore, we consider Complexity to be a relevant factor in adoption of Privacy-ABCs.

Trialability: Refers to the degree to which an innovation may be experimented with and tested on a limited basis [26]. In another word, it concerns how easy it would be for a potential adopter to test (or partially test) the features that the new technology provides. There is always a level of uncertainty for the adopters when they decide to invest on a new innovation and reducing this uncertainty by allowing them to try out the innovation would probably influence their decision. Trialability has been identified to be important in a number studies concerning

adoption of a new technology [16,18,24]. Borking [3] also highlights the role of Trialability for the adoption of PETs in general. Therefore we envision that Privacy-ABCs adoption can be influenced by their Trialability.

Observability: Refers to the extent to which the innovations are visible for the outside world [26]. In this regard, Moore and Benbasat [20] consider demonstrability as one type of observability. Unlike many other innovations that have visible results and can be well demonstrated, Privacy-ABCs are very challenging to present. They are not like standalone products and are always integrated into another service in order to perform access control. Therefore, demonstrators have difficulties showing all the added values of Privacy-ABCs in demos.

Perceived Benefits/Costs

It has been indicated in both DOI and Iacovou models that the likelihood of the allocation of the managerial, financial, and technological resources necessary to use that innovation are increased when there is better managerial understanding of the relative advantage of an innovation increases. Employing Privacy-ABCs comes with some direct and indirect costs such as for implementation, education, and change of processes. It can be challenging to find business cases that are enabled by Privacy-ABCs directly, nevertheless, compliance with data protection regulation, less investment in personal data storage and protection, and reduced risk of privacy breaches are the perceptions that can influence Privacy-ABCs adoption.

Organization

Beside the characteristics of an innovation itself, several organizational characteristics of the potential adopters have an influence on their decision to adopt or reject an innovation.

Top Management Support: It has been shown that technology innovation adoption can be influenced by top management support and their attitudes towards change [7,22,32]. The lack of top management support was identified as a key inhibitor in B2B deployment of e-commerce by [30]. Borking [3] also mentioned that top management's attitude towards changes caused by PETs can influence the adoption of PETs. Consequently, it may be the case that top management support increases the likelihood of adoption of Privacy-ABCs.

Business Dependency on User Data Collection: The role of industry sector in which a firm operates has been investigated and identified to have an influence on adoption of IT technologies in [11,19,25]. Indeed, Privacy-ABCs have not been invented only for the use of businesses and enterprises and can be adopted by any other organization. Nevertheless, we experienced in the context of the ABC4Trust EU research project that it is usually more challenging to convince businesses to integrated Privacy-ABCs into their services compared to other organizations such as non-profit ones. That increased the curiosity to have a

special look at the case of commercial adopters that might result in a useful impression of the influencing factors. Essentially, the *Business Model* of a company defines its roadmap. Osterwalder [21] defines the business model to be a conceptual link forming a triangle between strategy, business organization and ICT. Among the elements of a business model, employing Privacy-ABCs can influence the following:

- Product: Adopting Privacy-ABCs allows the customers to reveal less personal information. Therefore, if the business model value proposition is shaped around the users' data (such as Social Networks), the company will be probably more reluctant to employ such kind of technologies. As a result, we expect that dependency of the business to the collected users data plays an important role in the decision for adopting Privacy-ABCs.
- Customer Interface: It is a common practice to conduct targeted marketing and advertisement based on the extra information collected from the users, such as demographic data. Using Privacy-ABCs heavily influence this part as they prevent the service providers from such kind of data collection.

Technology Competence: Technological resources have been consistently identified as an important factor for successful information systems adoption [8,32]. A higher perceived technical competence was also identified by [17] as a key factor in adoption of electronic data interchange. The work by [36] also demonstrated that technology competence significantly drives e-business usage. So, the role of technology competence has been proven in the literature in adoption of many IT innovations. We consider technical competence to be relevant for adoption of Privacy-ABCs as we also experienced in the context of ABC4Trust pilots that typical developers had difficulties to integrate Privacy-ABCs into some services on their own and constant support of technology providers was needed, while developers with scientific background and technical understanding of the technology went through the integration process smoothly. Hence, lack of technical competency can hinder Privacy-ABCs adoption.

External Pressure

As we mentioned earlier, various sources of external pressure may influence the adoption of new innovations. Here we briefly introduce the ones that are relevant for Privacy-ABCs.

Regulatory Pressure: A regulatory body may be the source of coercive pressures [29]. In this regard, there has been movements in the regulatory sectors in Europe introducing more restriction on users' data collection and processing (Art. 6 and 7 of Directive 95/46/EC) as well as secure storage of the collected data (Art. 16 and 17 of Directive 95/46/EC). Consequently, it can be foreseen that organizations will soon feel pressure to start reconsidering their data collection schemes and look for secure solutions that reduces their liability for protecting the users data.

Social Pressure: There have been major incidences recently which we expect them to have an influence on adoption of privacy enhancing technologies in general. The most well-known incidence was brought up by Edward J. Snowden[4], which indeed highly stimulated the public opinion on the need for a raise of privacy in online environments. So, we expect that social pressure on the service providers will increase and therefore urges them towards employing mechanisms that reduce personal data collection in their processes.

Extent of Adoption among Competitors: The existence of mimetic pressures toward the adoption of innovations by organizations is confirmed in [10,13]. Knowing a competitor has adopted an innovation and it has been a success, the firm tends to adopt the same innovation [14]. The work by [29] confirmed the strong role of this factor in adoption of E-Procurement System. It could happen that offering more privacy becomes an advertising parameter especially in countries with more privacy protection culture. Therefore adoption of Privacy-ABCs by the competitors of a firm can motivate the decision makers to follow the same approach not to lose on the trust reputation.

Standardization: It is very typical for industries to employ procedures, processes or protocols that are standardized in order to ensure interoperability and sustainability of their products and services. In this regard, Standardization can become a source of normative isomorphism. There have been standardization projects that are very relevant to Privacy-ABCs and the ABC4Trust architecture [12]. For instance, ISO/IEC 24760 focuses on a framework for identity management and is conducted in 3 parts covering *Terminology and concepts*, *Reference architecture and requirements*, and *Practice*. Such standards have a good potential to influence the future adoption of Privacy-ABCs.

Environment

Here with environment we refer to the external conditions that do not introduce any pressure but can facilitate or hinder adoption of an innovation. For instance, it is more likely to succeed in implementing the idea of a remote movie rental company in a country that has cheaper, faster and more reliable postal services around.

Established Infrastructure Readiness: Electronic IDs have been implemented in various countries around the world, and therefore use of digital certificates for authentication and access control have been leveraged for service providers. Privacy-ABCs have been demonstrated their capabilities to be integrated with the existing eID infrastructure [2]. Furthermore, the European Commission also considered investing on the research for integration of Privacy-ABCs into future electronic IDs[5]. Consequently, having the global infrastructure ready to support Privacy-ABCs, the integration of these technologies into authentication and access control of service providers will be facilitated.

[4] http://en.wikipedia.org/wiki/Edward_Snowden.
[5] http://www.futureid.eu/.

4 Empirical Evaluation

4.1 Methodology

As Privacy-ABCs are not yet adopted, it is not possible to survey the service providers (adopters/non-adopters) in order to discover the drivers and the inhibitors. Thus, we decided to follow a forecast approach and collect the opinion of the experts from the relevant fields on the importance and influence level of the potential factors we introduced in our conceptual model. We designed a questionnaire containing quantitative and used a 5-point Liker scale from "not important at all" or "not at all influential" to "extremely important" or "extremely influential". Moreover, based on our experience of the ABC4Trust pilots, we made some of the factors more granular and presented them in two questions. That includes the Cost factor, which we presented as *Cost of Integration* and *Cost of Education*, Complexity factor, divided into *Complexity for Developers* and *Complexity for Users*, and the Government Pressure, presented as *Regulations for Data Collection* and *Regulations for Securing the Collected Personal Data*.

We refined the questionnaire in an iterative process performed in four steps with the help of two groups, one *with* dominant knowledge of Privacy-ABCs, and one *without* dominant knowledge of Privacy-ABCs. In the first step, a person with dominant knowledge in the field reviewed the questionnaire to check the technical correctness and readability of the questions. After rounds of discussions a version was ready to be reviewed by the people without dominant knowledge to validate the readability of the questions. After receiving their feedbacks, the questions were modified to improve the readability. In the next step, again a person with dominant knowledge reviewed the changes and the proposed updates. The next version was then distributed to the people without dominant knowledge and as we did not receive further clarification requests, the questionnaire was finalized. In this questionnaire, the respondents were asked to evaluate their level of expertise using the five-level Dreyfus model of skill acquisition. They were also requested to select their domains of expertise from relevant list including "Privacy and Identity Management", "Data Protection", "Policy Maker", and "Software and Services", or specify it if it was not on the list (multiple selection was allowed).

The survey was performed during the ABC4Trust summit event, on 20th of January 2015 in Brussels. The event was one of the best opportunities to get into contact with the experts of the relevant domains as it was broadly advertised via various important channels such as the one from the European Commission. Furthermore, having prestigious guest speakers also increased the chance of attracting stack-holders to the event. During this event, we gave a full day tutorial of Privacy-ABCs to the participants, covering various aspects such as limitation of current Identity Management Systems, how Privacy-ABCs work in theory, their implementation on computers, smartcards and mobile phones, as well as four real-time demos of some scenarios where Privacy-ABCs could improve users' privacy. These demos addressed a wide range of scenarios, namely "online university course evaluation system", "school community interaction platform",

"online movie streaming"[6], and "hotel booking"[7]. It is important to note that most of the tutorials and presentations were performed by the partners who were not involved in this study so we avoided unintentional biasing of the audience.

4.2 Results

At the end of the day, the participants were asked to answer the provided questionnaire. From over 80 participants, 20 completed the questionnaire, of which we excluded 3 as the respondents evaluated themselves below "Proficient" (below 4 out of 5). From the remaining respondents (the experts), 10 chose "Privacy and Identity Management", 3 chose "Data Protection", 3 chose "Policy Maker", and 5 chose "Software and Services" as their fields of expertise (multiple selection was allowed).

Table 1 summarizes the influence/importance level of the factors from the experts' perspective along with their ranking. The items ranked from 1 to 7 have a mean value over 3.0, meaning that the experts considered them on average "very" or "extremely" important or influential. The results show that in experts' opinion, "Technical Competency" of the adopters has the least effect on their decision among the others. Nevertheless, all the factors received a mean score over the average (2.0).

5 Implications and Discussion

From a practical point of view, the results give directions to supporting communities showing them where to put their future efforts. To foster adoption of Privacy-ABCs, priority shall be given to the items ranked from 1 to 7 ($mean > 3.0$). In this regard, the opinion of the experts can be reflected in two dimensions:

First, the Privacy-ABCs technology developers shall enrich the implementation of Privacy-ABCs in terms of

- Usability and Risk Communication: Privacy-ABCs such as different anonymity levels or applying predicates over attributes did not exist in the previous generation of access control mechanisms. Thus user interfaces shall be enhanced to appropriately communicate such features. In addition to that, Privacy-ABCs are user-centric approaches and their implementation essentially requires a piece of software to run on behalf of the users. This urges the users to install a client agent to represent them in the protocol steps, which consequently reduces the mobility of the users as they need to have this software on every device they use. In this regard, new deployment schemes reducing the need for client side installation can support reducing the complexity for the users.
- Agile Trial Platforms: Having online services that allow the interested parties to rapidly and with minimal effort integrate Privacy-ABCs for trial purposes can significantly improve their trialability.

[6] https://idemixdemo.mybluemix.net/.
[7] https://abc4trust.eu/demo/hotelbooking.

Table 1. Ranking of the relevant factors for adoption of Privacy-ABCs by service providers based on the experts' opinions

Rank	Factor	Mean	Var.
1	Business model dependency to data collection	3,71	0,22
2	Complexity for users	3,53	0,39
3	Observability	3,29	0,60
3	Top management support	3,29	1,47
5	Trialability	3,24	0,32
6	Cost of integration	3,19	0,83
7	Complexity for developers	3,12	0,74
7	Regulations for data collection	3,12	0,61
9	Regulations for securing the collected personal data	2,94	0,68
10	Established infrastructure readiness	2,88	0,99
11	Social pressure	2,71	1,10
12	Compatibility with existing IdM infrastructure	2,65	1,49
13	Competition among service providers	2,59	0,76
14	Cost of education	2,53	0,89
14	Standardization	2,53	1,26
16	Technical competency	2,35	1,12

- Designing Comprehensive Demos: In our questionnaire we asked the experts to select the most informative demos they saw during the day. The school community interaction platform received the most points (9 votes). The experts mainly mentioned they liked the fact that it was a complete set of scenarios and there were very many roles and a richer set of credentials. This allowed to show similarities and differences in the policies and implementations. However, the university course evaluation demo received the second highest point (6 votes) and the given reason was that the smaller scope made the scenario basic, very clear and easy to understand the benefits.
- Plug-and-Play Libraries: providing robust, rich and plug-and-play libraries along with appropriate documentation can notably facilitate the integration process for the software developers and consequently decrease the integration costs.

The second dimension relates to the dissemination strategies. More effort shall be put to target high-ranked managers and provide them with supporting materials that raise their understanding and awareness of Privacy-ABCs such as what these technologies can offer, how Privacy-ABCs can influence their processes, and what is needed for them to employ Privacy-ABCs. Moreover, the data protection bodies and the policy makers shall try to disseminate the capabilities of Privacy-ABCs to the regulatory authorities so that they become aware of the technical means to enforce minimal data collection regulations.

From a theoretical perspective, our results contribute to the existing theories by delivering a reduced conceptual model (Fig. 2) as a result of the expert surveys. Compared to the literature, our conceptual model introduces a new potential factor for adoption of technologies that limit service providers' access to users' data. The low variance clearly confirms that most of the experts had similar opinion on the role of "Business Model Dependency to Data Collection" and considered it as a key factor for shaping the desire of the service providers to adopt Privacy-ABCs.

The conceptual model also triggers theoretical research to boost the identified factors. More specifically, we see open questions on the methods to efficiently, transparently and explicitly communicate identity and attribute disclosure risks to the users via corresponding user interfaces of Privacy-ABCs.

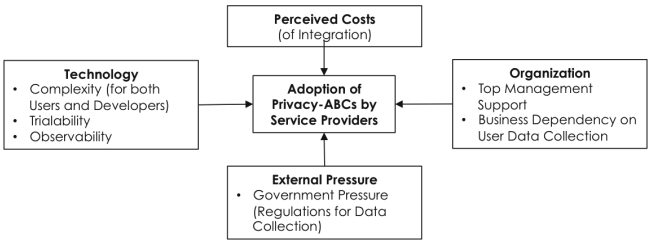


Fig. 2. Reduced conceptual model for factors influencing adoption of Privacy-ABCs by service providers

6 Conclusion

Privacy-ABC technologies are promising mechanisms that allow privacy and security go hand-in-hand. They provide various privacy features such as minimal attribute disclosure as well as unlinkable partial identities. Privacy-ABCs have passed the trial phase and proved their applicability and it is now important to understand how we can push these technologies forward. In this work, we investigated the potential factors that influence adoption of Privacy-ABCs by service provider and empirically evaluated the developed conceptual model through expert surveys. We collected the opinion of the experts during an especial international event where they received a full-day tutorial of various aspects of Privacy-ABCs.

The statistics of the collected opinions show that *Business Model Dependency to Data Collection*, *Complexity for User*, *Top Management Support*, *Observability*, *Trialability*, *Cost of Integration*, *Regulations for Data Collection*, and *Complexity for Developers* are the most important or influential factors impacting the decision of the service providers to employ Privacy-ABCs. These findings put lights on the directions towards which the supporting community should move and imply, despite the common beliefs of recognizing Privacy-ABCs as a redeemer to fight Social Networks, Privacy-ABCs may have higher chance to succeed in their adoption if they first target the service providers in the markets that are not based on users' data.

References

1. Sabouri, A. (ed.): Architecture for attribute-based credential technologies - final version. Deliverable D2.2, The ABC4Trust EU Project (2014). https://abc4trust.eu/download/Deliverable_D2.2.pdf. Accessed 08 November 2014
2. Bjones, Ronny, Krontiris, Ioannis, Paillier, Pascal, Rannenberg, Kai: Integrating Anonymous Credentials with eIDs for Privacy-Respecting Online Authentication. In: Preneel, Bart, Ikonomou, Demosthenes (eds.) APF 2012. LNCS, vol. 8319, pp. 111–124. Springer, Heidelberg (2014)
3. Borking, J.J.: Why adopting privacy enhancing technologies (pets) takes so much time. In: Gutwirth, S., Poullet, Y., de Hert, P., Leenes, R. (eds.) Computers, Privacy and Data Protection: An Element of Choice, pp. 309–341. Springer, Netherlands (2011)
4. Brands, S.: Untraceable off-line cash in wallet with observers. In: Helleseth, T. (ed.) Advances in Cryptology - CRYPTO 1993, vol. 765. Springer, Heidelberg (1994)
5. Camenisch, Jan L., Lysyanskaya, Anna: Signature Schemes and Anonymous Credentials from Bilinear Maps. In: Franklin, Matt (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 56–72. Springer, Heidelberg (2004)
6. Cavoukian, A., et al.: Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada (2009)
7. Chong, A.Y.L., Ooi, K.B., Lin, B., Raman, M.: Factors affecting the adoption level of C-Commerce: an empirical study. J. Comput. Inf. Syst. **50**(2), 13 (2009)
8. Crook, C.W., Kumar, R.L.: Electronic data interchange: a multi-industry investigation using grounded theory. Inf. Manag. **34**(2), 75–89 (1998)
9. DiMaggio, P.J., Powell, W.W.: The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. Am. Sociol. Rev. **48**(2), 147–160 (1983)
10. Fligstein, N.: The spread of the multidivisional form among large firms, 1919–1979. Adv. Strat. Manag. **17**, 55–78 (1985)
11. Goode, S., Stevens, K.: An analysis of the business characteristics of adopters and non-adopters of world wide web technology. Inf. Technol. Manag. **1**(1–2), 129–154 (2000)
12. Hansen, M., Obersteller, H., Rannenberg, K., Veseli, F.: Establishment and prospects of Privacy-ABCs. In: Rannenberg, K., Camenisch, J., Sabouri, A. (eds.) Attribute-based Credentials for Trust, pp. 345–360. Springer International Publishing, Switzerland (2015). http://dx.doi.org/10.1007/978-3-319-14439-9_11
13. Haunschild, P.R., Miner, A.S.: Modes of interorganizational imitation: the effects of outcome salience and uncertainty. Adm. Sci. Q. **42**(3), 472–500 (1997)
14. Haveman, H.A.: Follow the leader: mimetic isomorphism and entry into new markets. Adm. Sci. Q. **38**(4), 593–627 (1993)
15. Iacovou, C.L., Benbasat, I., Dexter, A.S.: Electronic data interchange and small organizations: adoption and impact of technology. MIS Q. **19**(4), 465–485 (1995)
16. Kendall, J.D., Tung, L.L., Chua, K.H., Ng, C.H.D., Tan, S.M.: Receptivity of singapore's SMEs to electronic commerce adoption. J. Strat. Inf. Syst. **10**(3), 223–242 (2001)
17. Kuan, K.K., Chau, P.Y.: A perception-based model for edi adoption in small businesses using a technology-organization-environment framework. Inf. Manag. **38**(8), 507–521 (2001)
18. Martins, C.B., Steil, A.V., Todesco, J.L.: Factors influencing the adoption of the internet as a teaching tool at foreign language schools. Comput. Edu. **42**(4), 353–374 (2004)

19. Miller, N.J., McLeod, H., Ob, K.Y.: Managing family businesses in small communities. J. Small Bus. Manag. **39**(1), 73–87 (2001)
20. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf. Syst. Res. **2**(3), 192–222 (1991)
21. Osterwalder, A., et al.: The business model ontology: a proposition in a design science approach (2004)
22. Premkumar, G., Ramamurthy, K., Crum, M.: Determinants of EDI adoption in the transportation industry. Eur. J. Inf. Syst. **6**(2), 107–121 (1997)
23. Premkumar, G., Roberts, M.: Adoption of new information technologies in rural small businesses. Omega **27**(4), 467–484 (1999)
24. Ramdani, B., Kawalek, P.: SME adoption of enterprise systems in the Northwest of England. In: McMaster, T., Wastell, D., Ferneley, E., DeGross, J. (eds.) Organizational Dynamics of Technology-Based Innovation: Diversifying the Research Agenda, IFIP International Federation for Information Processing, vol. 235, pp. 409–429. Springer, USA (2007)
25. Raymond, L.: Determinants of web site implementation in small businesses. Internet Res. **11**(5), 411–424 (2001)
26. Rogers, E.: Diffusion of innovations. Free Press, New York (2003)
27. Sabouri, A., Krontiris, I., Rannenberg, K.: Attribute-Based Credentials for Trust (ABC4Trust). In: Fischer-Hübner, S., Katsikas, S., Quirchmayr, G. (eds.) TrustBus 2012. LNCS, vol. 7449, pp. 218–219. Springer, Heidelberg (2012)
28. Scott, W.R.: Institutions and Organizations (Foundations for Organizational Science). SAGE Publications Inc., New York (2000)
29. Soares-Aguiar, A.: Palma-dos Reis, A.: Why do firms adopt E-Procurement systems? using logistic regression to empirically test a conceptual model. IEEE Trans. Eng. Manag. **55**(1), 120–133 (2008)
30. Teo, T.S., Ranganathan, C., Dhaliwal, J.: Key dimensions of inhibitors for the deployment of web-based business-to-business electronic commerce. IEEE Trans. Eng. Manag. **53**(3), 395–411 (2006)
31. Teo, T.S., Tan, M., Buk, W.K.: A contingency model of internet adoption in Singapore. Int. J. Electron. Commer. **2**(2), 95–118 (1997)
32. Thong, J.Y.: An integrated model of information systems adoption in small businesses. J. Manag. Inf. Syst. **15**(4), 187–214 (1999)
33. Tornatzky, L.G., Fleischer, M., Chakrabarti, A.K.: Processes of technological innovation (1990)
34. Wästlund, E., Angulo, J., Fischer-Hübner, S.: Evoking Comprehensive Mental Models of Anonymous Credentials. In: Camenisch, J., Kesdogan, D. (eds.) iNetSec 2011. LNCS, vol. 7039, pp. 1–14. Springer, Heidelberg (2012)
35. Zafarani, R., Liu, H.: Connecting users across social media sites: a behavioral-modeling approach. In: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, 11–14 August 2013, pp. 41–49 (2013)
36. Zhu, K., Dong, S., Xu, S.X., Kraemer, K.L.: Innovation diffusion in global contexts: determinants of post-adoption digital transformation of european companies. Eur. J. Inf. Syst. **15**(6), 601–616 (2006)

Consumers' Perceptions of Social Commerce Adoption in Saudi Arabia

Salma S. Abed, Yogesh K. Dwivedi(✉), and Michael D. Williams

Management and Systems Section (MASS),
School of Management, Swansea University, Swansea, UK
{717185,y.k.dwivedi,m.d.williams}@swansea.ac.uk

Abstract. This study aims to examine the factors that affect consumer adoption of social commerce technologies in the context of Saudi Arabia. The factors descriptively explored in this research include: performance expectancy, effort expectancy, social influence, hedonic motivation, habit, trust, consumer innovativeness, information quality and behavioural intention. The survey data utilised in this research was collected through a self-administered questionnaire within a convenience sample. The results obtained through a descriptive analysis confirmed that Saudi consumers perceive the abovementioned factors as important and they have strong behavioural intention to use social commerce technologies.

Keywords: Social media · Social commerce · Customer · UTAUT2 · Adoption · Saudi Arabia

1 Introduction

Social media plays an important part in the countries' economic development [23] as it offers novel ways for both organisations and customers to link with each other. Businesses began to embrace social media websites as a technique to improve communication, information sharing, and collaboration by applying many innovative and vital business practices [34]. Social media motivated companies to work faster by creating and operating more interdependencies in global markets [1]. Therefore, the development of social media has enhanced a new e-commerce model named social commerce. The term of social commerce is defined as a concept of Internet-based social media that allows people to contribute actively in the selling and marketing of diverse products and services in online marketplaces [20]. This dynamic process assists consumers to get better information about different products and services delivered by companies [15].

Saudi Arabia has witnessed the biggest growth of diffusion of social media platforms, which is the strongest supporting factor to e-commerce adoption within the country [9]. Many large businesses, as well as small companies and new ventures, have set-up their organisations and group profiles on Facebook, LinkedIn and other similar websites. In fact, Saudi Arabia ranks the second in the Arab countries, after Egypt, for registering 5,240,720 Facebook users [5]. In addition, Saudi Arabia represents the largest percentage of Twitter users with around 830,300 users or 38 % of total Arab

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 133–143, 2015.
DOI: 10.1007/978-3-319-25013-7_11

users [5]. Moreover, statistics have revealed that Saudi Arabia has contributed 90 million video views per day. This is the maximum number of YouTube viewings worldwide per Internet user [9]. These amazing evidences have created a new background for business owners, managers, and marketers to reach their potential consumers. However, even though consumers in other countries such as Hong Kong, China, South Korea, and Thailand apply online shopping activities actively, mostly Saudi consumers use online social media simply to help them make buying decisions [3]. The remaining sections of the paper include the relevant literature review and theoretical basis in the third section. These are followed by the methodology, results, and discussion outlined in the fourth, fifth and sixth sections, respectively. Finally, the research contribution, conclusion and limitations are delivered in section seven eight, and nine.

2 Literature Review

Limited empirical studies have been conducted within this context of social commerce as it is still new. [14, 15] extended the Technology Acceptance Model (TAM) to measure social commerce adoption by consumers. The researcher examined some components of social commerce that affected consumers' intention to buy. The model tested ratings and reviews, referrals and recommendations, and forums and communities aiming to help to introduce new business plans for e-vendors. The study also indicated that trust is a continuing issue in e-commerce and can be examined in social commerce constructs. The researcher collected survey data and applied structural equation modelling (SEM) for analysis. The results pointed out that perceived usefulness and forums and communities have a positive impact on trust. In addition, the findings showed that trust has a significant effect on intention to buy [14, 15].

In the context of Saudi Arabia, Only two studies have been found that focused on the use of online social media to help overcome e-commerce adoption barriers. [4] presented research that is designed to demonstrate a conceptual framework by extending it from TAM. The framework aims to examine social media effects and perceived risk as the moderating effects between purchasing intention and an actual online purchase in Saudi Arabia. Furthermore, [17] evaluated the effectiveness of online social networking by entrepreneurs in the Arabian Gulf including Bahrain, Kuwait, United Arab Emirates, Saudi Arabia, Qatar, and Oman. The research used a qualitative approach by interviewing a sample size of 50 business entrepreneurs in the Arabian Gulf who used online social networks as a method of promoting their products. The study found that online social networks are a cheap and easy method of advertising and would give all entrepreneurs a better chance of reaching their target market as well as succeeding in their ventures. In addition, entrepreneurs can now target their markets using online social networks. Moreover, social networking websites allow businesses to introduce their products to different market segments, with a low chance of failure and low expense. Finally, the researchers added that the rise in web-based social interaction can change the way businesses operate in the future. Finally, empirical studies conducted in the context of social commerce are very limited, which this study aims to address. This study proposes a new model that can be

developed and extended by applying more recent and comprehensive technology adoption theories as well as adding other more appropriate constructs especially in the consumer context.

3 Theoretical Basis

In order to explain social commerce adoption effectively from the customers' perspective, the conceptual model should provide a clear image of social commerce features. From the analysis of the common theories in the field of technology acceptance, [31, 33] developed the Unified Theory of Acceptance and Use of Technology (UTAUT) by merging eight IT acceptance models. The UTAUT has four main constructs including performance expectancy, effort expectancy, social influence, and facilitating conditions that influence behavioural intention to use a technology and usage behaviours. UTAUT was capable to explain around 70 per cent of the variance of behavioural intention and around 50 per cent of the variance of use behaviour [31, 33]. Recently, [32] proposed an additional three constructs to the original UTAUT model including hedonic motivation, price value, and habit. [32] claimed that the suggested additions to UTAUT2 present major changes in the variance described in behavioural intention and use behaviour especially within the consumers' context.

Considering the limits of the earlier investigated constructs for social commerce adoption [14, 15], UTAUT2 is more suitable in the context of this study. This is because UTAUT2 was developed on UTAUT, which has been credited as the most parsimonious and comprehensive predictive model [31, 32]. Furthermore, UTAUT2 is proposed mainly for explaining technology acceptance from the customers' contexts other than the organisational use [32]. Besides, the UTAUT2 has investigated factors influencing users' acceptance of iPad phones [16], mobile payment uses [29], and mobile learning acceptance [19] that shares similar technological characteristics with social commerce. As a result, the UTAUT2 has been selected as the theoretical foundation of the proposed conceptual model in order to understand the antecedents of customers and their preference to buy from social media websites, in a new cultural context (Saudi Arabia). This shadows [32] the suggestion that future research should apply UTAUT2 in diverse countries. This also follows the call by [32] for future research to investigate the UTAUT2 on different technologies. Furthermore, to be consistent with the recommendations of [32], other external factors (trust, consumer innovativeness, and information quality) will be measured along with UTAUT2 constructs (performance expectancy, effort expectancy, social influence, hedonic motivation, habit) in the same conceptual model. These other constructs have been developed according to the literature review.

Furthermore, discard the two independent variables facilitating condition and price value and the dependent variable use behaviour which are considered a part from UTAUT2 model's constructs. Discarding both independent variables is a logical step as social commerce use does not need any technical infrastructure support as in system adoption; apart from the Internet cost, there is no monetary cost for using social media because most of the time, wireless connections are available free of charge. Thus, discarding the dependent variable use behaviour is because the literature indicated that

social commerce technologies are new in the Arabian Gulf generally and in Saudi Arabia specifically and they are still not fully used [2]. The suggested constructs are demonstrated below in Table 1:

Table 1. Description of factors

Constructs examined	Definition
Performance expectancy	"The degree to which an individual believes that applying the technology will help him or her to gain in job performance" [31].
Effort expectancy	"Extent of ease connected with the use of system" [31].
Social influences	"The extent to which an individual perceives that important others believe he or she should apply the new system" [31].
Hedonic motivation	"The feeling of cheerfulness, joy, and enjoyment, which is stimulated by applying technology" [32].
Habit	"The extent to which, people tend to perform behaviours automatically because of learning" [32].
Trust	"Individual willingness to depend based on the beliefs in ability, benevolence, and integrity" [12]
Consumers' innovativeness	"The degree to which the individual is willing to adopt innovations such as goods and services or new ideas without communicating with others' previous purchasing experience" [24]
Information quality	"The consumers' general perception of the accuracy and completeness of website information as it relates to products and transactions" [21].
Behavioural intention	"The extent to which an individual intends to adopt the technology in the future" [31].

4 Methodology

This study utilised a total of thirty-six scale items that were derived from the literature of technology adoption in order to measure the selected variables; i.e. Trust, consumers innovativeness and information quality in addition to other variables in the UTAUT2 model. To do that, it is vital to make items that will help in measuring the characters of the related variants. Several items were used in order to measure the variables of the UTAUT2 which were adapted from [32]. The additional variables have used items that were adopted by a number of writers. Therefore, trust was selected from [12, 18, 21, 26]. In addition, this paper selected the consumer innovativeness items from [27]. Furthermore, information quality items have been adopted from [21]. The degree of responses was estimated using the seven-point scale ranging from strongly agree to strongly disagree. In regards to the language of the data collection tool, the questionnaire was translated into Arabic to overcome the cultural and linguistic differences [6]. Then, a pilot study was conducted using 20 questionnaires that were distributed to Saudi social media users who were asked to give their feedback in case faced any difficulties in answering the questionnaire [10]. Accordingly, the questionnaire's items were rechecked in terms of clarity, language simplicity, and length.

When it came to the sampling, this study implemented a convenience sampling as the researcher does not have a list of social commerce potential users. Additionally, convenience sampling is cost-effective [10, 11]. Furthermore, the results of a convenience sample can be generalised more appropriately since it allowed for the presence of a variety of profiles and characters of potential users [11]. The population that was going to be sampled were all from the regions of Saudi Arabia including big cities and small towns. This has been achieved by distributing the survey questionnaires with both hard copies and online as web links. In the hard copy survey, most of the respondents of the questionnaires were students, as the questionnaires were distributed to distance-learning students in the Management School at King Abdul-Aziz University in Jeddah. The students were present for two weeks at the university campus for the final exams. The researcher took the chance to distribute the survey to students. Distance-learning students enrolled in the programme are from different regions of the Kingdom including large cities and small towns; there were different age groups; and from both genders. Therefore, they were representative of the diverse population of Saudi Arabia. In the soft copy of the survey, the researcher used the online survey software Qualtrics for distributing the web-based survey. The web link was sent to different Saudi Arabian e-commerce groups on social media such as social networking sites including Facebook and LinkedIn; as well as the micro blogging service such as Twitter. Due to space constraints scale items/measurements cannot be provided in the paper but will be available upon the request.

5 Results

5.1 The Response Rates

As mentioned earlier, the survey questionnaires were circulated in both hard copies and online as web links. In the hard copy survey, participation was completely voluntary. Respondents were requested to complete a questionnaire based on their perception and/or acceptance of social commerce. A total of 700 survey questionnaires were distributed and the returned completed surveys were 417 with a 59 per cent response rate. In the soft copy of the survey, the researcher used the online survey software Qualtrics for distributing the web-based survey. One of the features the online survey software Qualtrics provided was that it showed how many respondents started the survey, but did not complete it. The total number of consumers who participated in the survey was 225. The total number of consumers who completed the survey was 120, with a response rate of 53 per cent. As a result, a total of 537 survey questionnaires were collected from both the paper-based survey and web-based survey. The questionnaires were carefully checked before entering the data using SPSS 22.0. Out of the 537 questionnaires collected, only 507 were used; 27 were considered unusable and discarded due to the huge amount of missing data resulting from missing pages or incomplete sections. In addition, three questionnaires were also discarded due to them having the same answer to all questions. These responses were considered as invalid, and they were removed in the data-editing process stage [28].

5.2 Respondents' Profile and Characteristics

When it comes to respondents' profile and characteristics, this research adopts the following demographic information: gender, age, and education (see Table 2). The demographic details of the main survey sample show that the majority of the respondents were female, forming 65.1 % of the whole sample, while males are represented by only 34.9 %. In regard to the respondents' age, the descriptive statistics demonstrate that the largest age population was within 21–29 years old with 62.7 %, followed by the age group of > = 18–20 with 17.9 %. The rest of the percentages were divided among the age group of 30–39 (13.4 %) and 5.5 % for those who were between 40–49, whereas the smallest percentage was 0.4 % as only two respondents were at the age of 50 and above. Regarding the educational level, the majority of respondents hold a Bachelor's degree, representing 45.0 % of the total sample. The second largest group were high school holders (38.9 %) followed by 10.5 % as postgraduates and 5.5 % as diploma holders. A very small percentage of respondents held less than high school qualification with (0.2 %). Table 2 shows the demographic details of the respondents in the main survey sample.

5.3 Descriptive Analysis and Normality Assumption

According to Table 3 below, the descriptive statistics show that there are three items devoted to measure consumers' perceptions on performance expectancy (PE). PE2 achieved the highest mean score of 5.61 (±1.302). In contrast, the lowest mean was 5.36 (±1.436) as a value recorded for PE3. There are four items identified on effort expectancy (EE). As seen, the largest mean scores were 5.56 (±.1.430) for EE3 and the lowest mean is for EE2, 5.24 (±.1.445). Social influence (SI) was measured by three

Table 2. Respondents' profile and characteristics

Variable	Group	Frequency	Percent
Gender	Male	177	34.9
	Female	330	65.1
	Total	**507**	**100.0**
Age	> = 18–20	91	17.9
	21–29	318	62.7
	30–39	68	13.4
	40–49	28	5.5
	50 and above	2	0.4
	Total	**507**	**100.0**
Education	Less than high school	1	0.2
	high school	197	38.9
	Diploma	28	5.5
	Bachelor's degree	228	45.0
	Postgraduate	53	10.5
	Total	**507**	**100.0**

Table 3. Descriptive and normality tests

Constructs	Descriptive statistics							
	Items	N	Mean	Std. deviation	Skewness		Kurtosis	
					Statistic	Std. error	Statistic	Std. error
Performance expectancy (PE)	PE1	507	5.56	1.333	–.861	.108	.583	.217
	PE2	507	5.61	1.302	–1.126	.108	1.389	.217
	PE3	507	5.36	1.436	–.707	.108	–.032	.217
Effort expectancy (EE)	EE1	507	5.46	1.440	–.891	.108	.468	.217
	EE2	507	5.24	1.445	–.769	.108	.210	.217
	EE3	507	5.56	1.430	–1.137	.108	1.091	.217
	EE4	507	5.33	1.539	–.936	.108	.303	.217
Social influence (SI)	SI1	507	4.47	1.678	–.428	.108	–.417	.217
	SI2	507	4.79	1.600	–.443	.108	–.437	.217
	SI3	507	5.09	1.487	–.652	.108	–.048	.217
Hedonic motivation (HM)	HM1	507	5.55	1.505	–1.121	.108	.887	.217
	HM2	507	5.63	1.382	–1.109	.108	1.113	.217
	HM3	507	5.56	1.422	–.857	.108	.144	.217
Habit (HT)	HT1	507	4.65	1.720	–.377	.108	–.615	.217
	HT2	507	4.15	1.892	–.191	.108	–.956	.217
	HT3	507	4.79	1.536	–.341	.108	–.727	.217
	HT4	507	4.71	1.776	–.514	.108	–.598	.217
Trust (TR)	TR1	507	4.33	1.826	–.293	.108	–.855	.217
	TR2	507	4.45	1.652	–.287	.108	–.577	.217
	TR3	507	4.37	1.737	–.288	.108	–.766	.217
	TR4	507	4.38	1.612	–.260	.108	–.590	.217
	TR5	507	4.91	1.536	–.550	.108	–.162	.217
Consumer innovativeness (CI)	CI1	507	5.16	1.542	–.672	.108	–.038	.217
	CI2	507	5.39	1.484	–.920	.108	.532	.217
	CI3	507	5.48	1.521	–1.010	.108	.501	.217
	CI4	507	4.94	1.688	–.608	.108	–.366	.217
	CI5	507	4.71	1.584	–.384	.108	–.409	.217
	CI6	507	4.66	1.679	–.377	.108	–.561	.217
Information quality (IQ)	IQ1	507	4.92	1.496	-.324	.108	–.446	.217
	IQ2	507	5.41	1.327	–.596	.108	–.058	.217
	IQ3	507	4.72	1.568	–.435	.108	–.266	.217
	IQ4	507	4.91	1.461	–.415	.108	–.215	.217
	IQ5	507	4.81	1.435	–.309	.108	–.367	.217
Behavioural intention (BI)	BI1	507	5.28	1.479	–.641	.108	–.060	.217
	BI2	507	5.05	1.534	–.551	.108	–.227	.217
	BI3	507	5.12	1.552	–.577	.108	–.315	.217
Valid N (listwise) 507								

items. The highest mean scores were 5.09 (±1.487) for SI3 and 4.79 (±.1.600) for SI2 followed by 4.47 (±1.678) for SI1 as the lowest mean. Moreover, there are three items allocated to measuring consumers' perceptions on hedonic motivation (HM). The highest mean value is 5.63 (±1.382) recorded for HM2, while the lowest mean value is 5.55 (±1.505) recorded for HM1. Habit (HT) was measured by four items with 4.79 as the highest score recorded for HT3 and the lowest mean value is 4.15 for HT2. Table 3 also shows that there are five items identified to measure consumers' perceptions on trust (TR). TR5 had the largest mean value of 4.91 (±1.536) compared to TR1 that had the lowest mean score of 4.33 (±1.826). In addition, six items were identified to measure consumers' perceptions on consumer innovativeness (CI). CI3 recorded the highest value with 5.48 (±1.521). In contrast, the lower score recorded was for CI6, 4.66 (±1.679). Finally, three items were adopted to measure the behavioural intention (BI) construct. The lowest mean was for BI2 with a value of 5.05 (±1.534) while the highest mean score was 5.28 (±1.479) for BI1.

Screening the data for assessing the variables normality is a crucial step in the analysis [13, 22, 30]. Normality means the shape of normal distribution of metric variable and its correspondence [13]. Normality of a single variable can be measured statistically or graphically [7, 25, 30]. The failure to achieve normality can result from invalid statistical tests. This study has tested skewness and kurtosis at the item level. Skewness refers to the symmetry of distribution; the test indicates if the distribution is shifted or unbalanced to one side [30]. There are two types of skewness: positive skewness, when the distribution is shifted to the left; and negative skewness when it is shifted to the right [13]. Kurtosis refers to the peakness of distribution [30]. Peaked distributions are termed leptokurtic, whereas, flatter distributions are termed platykurtic. The values of skewness and kurtosis are zeroes when variables have normal distributions. Consequently, positive or negative values indicate a deviation from normality. The range of values for suitable deviations is affected by sample size. In small samples less than 30, slight deviations can be serious, whereas with large sample sizes with more than 200 it can be ignorable [13]. On the other hand, the most generally acceptable critical value for kurtosis and skewness distribution is ± 2.58 [13]. Table 3 indicates that skewness and kurtosis variables fall within the acceptable range.

6 Discussion

The presented descriptive results in this study help to visualise what the data revealed. After overviewing the literature review of the studies that have used IS theories to examine social commerce adoption, only one study has used the TAM model [14, 15]. As a result, this study adopted the UTAUT2 variables as a more recent and comprehensive technology adoption theory. Other more suitable constructs were added especially in consumer context (trust, consumer innovativeness, and information quality). Furthermore, this study has identified items to test the proposed variables, which has been examined in previous research in the literature. By collecting empirical data from 700 paper-based survey participants and 225 web-based survey participants, the study identified 507 valid participants response. The findings provided a summary

regarding the response rate, respondents' profile and characteristics, and a descriptive analysis and normality tests.

Regarding the descriptive analysis of the measurement items, the standard deviation is used to quantify the amount of variation or dispersion of the examined data values. When the standard deviation is close to 0, it indicates that the data points lean very close to the mean of the examined data values, but a high standard deviation shows that the data is spread over a wider range of values [8]. In other words, the low value of standard deviation reflects that there is a high certainty that most of the participants have similar views towards the variable. In this study, the average mean and standard deviation of all examined variables were in their recommended level. In addition to the normality test, Table 3 indicates that skewness and kurtosis variables fall within the acceptable range. As a result, it seems that the items of PE, EE, SI, HM, HT, TR, CI, IQ and BI were able to capture a high average mean with a suitable normality test results. Accordingly, it is worth stating that the majority of the survey questionnaire's respondents positively perceive the aspects associated to these constructs. Therefore, future research should take this into consideration so that the significance level of the dependent constructs over the behaviour intention to use social commerce can be examined; this will certainly guide the Saudi organisations to give more consideration towards the most significant factors that affects consumers to use social commerce technologies.

6.1 Research Contribution

The current study makes a significant contribution by proposing the UTAUT2 model for examining the adoption of social commerce technologies, which is a novel modern technology. Furthermore, the study also expanded the applicability of UTAUT2 by focusing on a new cultural context (that is: Saudi Arabia). Finally, this study is able to extend the theoretical horizon of UTAUT2 by including other external factors from the technology adoption literature.

7 Conclusion

This study aims to identify the important factors that influence the adoption of social commerce by Saudi customers. UTAUT2 has been identified as a suitable theoretical foundation for proposing a conceptual model. The study has added other significant and frequently used factors (trust, consumer innovativeness, and information quality) along with UTAUT2 constructs to formulate the model. In order to achieve the study's objectives, a quantitative field survey was conducted to obtain data from a convenience sample of Saudi customers; the data collection used a self-administered questionnaire. Finally, the researcher did a descriptive analysis for each one of the investigated variables. The findings indicated that these factors play a significant role in the behavioural intention for the participants.

7.1 Limitations and Future Research Directions

This study aimed at investigating behaviour intention to use social commerce technologies in the context of Saudi Arabia. First, conducting a descriptive analysis instead of inferential analysis will not allow the extension of the findings to the whole of the Saudi Arabian population. Consequently, this study will guide to assume the hypotheses in regards to the relations between factors, as well as using the structural equation modelling (SEM) to test the measurement model, structural model, and model fitness. Second, to consider investigating the behaviour intention rather than the actual use of social commerce will not give an overall view about using this technology in Saudi Arabia. Therefore, these issues should be taken into consideration in future research. This may assist organisations in Saudi Arabia to select the best strategy for encouraging consumers to use social commerce technologies, which will benefit their businesses. Finally, a comparative research should be conducted between developing and developed countries; also, the cultural context should be taken into consideration in the comparative research.

References

1. Abed, S., Dwivedi, Y.K., Williams, M.D.: Social media as a bridge to E-Commerce adoption in SMEs: a systematic literature review. Forthcoming Mark. Rev. **15**(1), 39 (2015)
2. Abed, S., Dwivedi, Y.K., Williams, M.D.: SMEs' adoption of E-Commerce using social media in a Saudi Arabian context: a systematic literature review. Int. J. Bus. Inf. Syst. **19**(2), 159–179 (2015)
3. ALMowalad, A., Putit, L.: Factors influencing saudi women consumer behavior in online purchase. J. Emerg. Econ. Islamic Res. (JEEIR) **1**(2), 1–13 (2013)
4. Al-Mowalad, A., Putit, L.: The extension of TAM: the effects of social media and perceived risk in online purchase. In: 2012 International Conference on Innovation Management and Technology Research (ICIMTR), pp. 188–192. IEEE, May 2012
5. Arab ICT Use and Social Networks Adoption Report.: Madar Research & Development, KACST (2012). http://www.kacst.edu.sa/en/about/publications/Other%20Publications/Arab%20ICT%20Use%20Report%202012.pdf. Accessed 05 January 2015
6. Brislin, R.: Comparative research methodology: cross-cultural studies. Int. J. Psychol. **11**(3), 215–229 (1976)
7. Coakes, S.J., Steed, L., Ong, C.: SPSS Version 16.0 for Windows. Wiley, Australia (2009)
8. Dancey, C.P., Reidy, J.: Statistics without Maths for Psychology. Pearson Education, Harlow (2007)
9. De Kerros Boudkov Orloff, A.: Ecommerce in Saudi Arabia: driving the evolution adaptation and growth of ecommerce in the retail industry. Sacha Orloff Consulting Group, SOCG, 17 June 2012. Accessed http://www.scribd.com/doc/136654512/E-Commerce-in-Saudi-Arabia-Driving-the-Evolution-Adaptation-and-Growth-of-Ecommerce-in-the-Retail-Industry-SOCG-2012June17
10. Dwivedi, Y.K., Choudrie, J., Brinkman, W.P.: Development of a survey instrument to examine consumer adoption of broadband. Indus. Manag. Data Syst. **106**(5), 700–718 (2006)
11. Franzosi, R.: From Words to Numbers: Narrative, Data, and Social Science, vol. 22. Cambridge University Press, Cambridge (2004)

12. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: an integrated model. MIS Q. **27**(1), 51–90 (2003)
13. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis, 7th edn. Pearson, Upper Saddle River (2010)
14. Hajli, M.: Social commerce adoption model. In: UK Academy for Information Systems Conference Proceedings 2012, Paper 16 (2012). http://aisel.aisnet.org/ukais2012/16
15. Hajli, M.N.: Social commerce for innovation. Int. J. Innov. Manag. **18**(4), 1–24 (2014)
16. Huang, C.Y., Kao, Y.S., Wu, M.J., Tzeng, G.H.: Deriving factors influencing the acceptance of pad phones by using the DNP based UTAUT2 framework. In: Technology Management in the IT-Driven Services (PICMET), 2013 Proceedings of PICMET 2013, pp. 880–887. IEEE (2013)
17. Indrupati, J., Henari, T.: Entrepreneurial success using online social networking: evaluation. Edu. Bus. Soc.: Contemp. Middle East. Issues **5**(1), 47–62 (2012)
18. Jarvenpaa, S.L., Tractinsky, N., Saarinen, L.: Consumer trust in an internet store: a cross-cultural validation. J. Comput.-Mediated Commun. **5**(2), 0 (1999)
19. Kang, M., Liew, B.Y.T., Lim, H., Jang, J., Lee, S.: Investigating the determinants of mobile learning acceptance in Korea using UTAUT2G. In: Chen, G., Kumar, V., Kinshuk, Huang, R., Kong, S.C. (eds.) Emerging Issues in Smart Learning, pp. 209–216. Springer, Heidelberg (2015)
20. Kim, S., Park, H.: Effects of various characteristics of social commerce (S-Commerce) on consumers' trust and trust performance. Int. J. Inf. Manag. **33**(2), 318–332 (2013)
21. Kim, D.J., Ferrin, D.L., Rao, H.R.: A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. Decis. Support Syst. **44**(2), 544–564 (2008)
22. Kline, R.B.: Principles and Practice of Structural Equation Modelling, 2nd edn. The Guilford Press, New York (2004)
23. Lea, B.R., Yu, W.B., Maguluru, N., Nichols, M.: Enhancing business networks using social network-based virtual communities. Indus. Manag. Data Syst. **106**(1), 121–138 (2006)
24. Midgley, D.F., Dowling, G.R.: Innovativeness: the concept and its measurement. J. Consum. Res. **4**(4), 229–242 (1978)
25. Pallant, J.: SPSS Survival Manual. McGraw-Hill, New York (2010)
26. Pavlou, P.A.: Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int. J. Electron. Commer. **7**(3), 101–134 (2003)
27. Roehrich, G.: Consumer innovativeness: concepts and measurements. J. Bus. Res. **57**(6), 671–677 (2004)
28. Sekaran, U.: Research Methods for Business: A Skill Building Approach, 4th edn. Wiley, New York (2003)
29. Slade, E.L., Williams, M.D., Dwivedi, Y.K.: Devising a research model to examine adoption of mobile payments: an extension of UTAUT2. Mark. Rev. **14**(3), 310–335 (2014)
30. Tabachnick, B.G., Fidell, L.S.: Using Multivariate Statistics. Pearson, Boston (2006)
31. Venkatesh, V., Morris, M., Davis, G., Davis, F.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
32. Venkatesh, V., Thong, J., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**(1), 157–178 (2012)
33. Williams, M.D., Rana, N.P., Dwivedi, Y.K.: The unified theory of acceptance and use of technology (UTAUT): a literature review. J. Enterp. Inf. Manag. **28**(3), 443–488 (2015)
34. Yates, D., Paquette, S.: Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. Int. J. Inf. Manag. **31**(1), 6–13 (2011)

Big and Open Data

Linking Operational Business Intelligence with Value-Based Business Requirements

Tom Hänel(✉) and Carsten Felden

Department of Management and Information Systems,
TU Bergakademie Freiberg, Freiberg, Germany
{tom.haenel, carsten.felden}@bwl.tu-freiberg.de

Abstract. Operational business intelligence (OpBI) integrates data of business processes to analyse their performance in relation to organizational goals. The consequent decision-making concerns a timely recognition and execution of actions to maintain performant business processes. OpBI systems can be designed according to a firm-specific definition of requirements guided by considerations from business model, business process and information system perspective. However, there is no approach to link the design of OpBI jointly with characteristics of business models and business processes, yet. The paper uses therefore an action research method and proposes a business approach that combines e^{3}value with the work system framework to set up conceptual application designs for an OpBI-reliant decision support. We report on results of a long-term research project to demonstrate the development and application of our approach in four different business scenarios. The findings include implications towards a business-oriented application design of OpBI systems.

Keywords: Operational Business Intelligence · e^{3}value · work system · ADAPT

1 Introduction

Organizations measure business processes using performance indicators in terms of time, quality, or cost [1]. The maintenance of performant business processes has to be closely linked to business strategy so that process improvements are valuable and lead to competitive advantages [2]. Management activities of process performance are thereby associated with IT to collect and analyse data about business processes [3]. Such IT capabilities need to be correspondent and compatible to business strategy, too, to avoid missing of expected performance results [2]. One possible concept to analyse business processes is OpBI dealing with an integration of daily business data [4]. This supports business operation's managers in gaining relevant knowledge to evaluate business process performances [5]. Management actions taken in consequence of an OpBI-reliant decision-making have to bring benefits to the manner of an organization creating value in its business environment. The paper's goal is therefore to investigate a linkage of OpBI with firm-specific business requirements.

The current discussion about OpBI provides no conceptual insights to consider business requirements for designing analytical systems in a particular case. For

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 147–159, 2015.
DOI: 10.1007/978-3-319-25013-7_12

instance, the analysis requirements of insurance companies differ from issues of automotive suppliers from business perspective, although the technical system components can be quite similar. It is not obvious for application developers, how an OpBI system needs to be logically designed in order to maintain and improve performant business processes from a perspective of business operation's managers. A specification of OpBI systems can benefit from a value-based requirements engineering so that business value models initialize requirements for business processes and IT systems [6]. We investigate such a value-based requirements definition for OpBI systems and propose an approach to link the logical design of analytical databases with firm-specific business value models. The paper contributes with a development and application of our approach to the scientific discussion using participatory action research in context of four different organizations. This offers collaborative insights for research and practice to the discourse about business approaches so that operational management actions are beneficial for performant business processes.

Chapter 2 refines the problem of research and analyses related areas. The research method is presented in Chap. 3. Chapter 4 introduces our approach and Chap. 5 reports on its application during an action research project in four different business scenarios. Finally, a conclusion summarizes findings and further research activities.

2 Status quo

OpBI is understood as a decision support concept for business operation's managers to analyse business processes in favour of continuous improvements of process design and execution [4]. OpBI supports an identification of control actions based on timely relations between process performance and the status of goal achievement. [5]

2.1 Problem Refinement

OpBI integrates data emerging in or flowing into IT systems during operational task fulfilment [4, 5]. From a technical viewpoint, OpBI systems can be equipped with IT providing business operation's managers access to manifold sources of information and analytical options in combination with high performance data processing. The discussion about Hadoop [7], cloud computing [8], combinations of transactional and analytical databases [9], or data virtualization techniques [10] points to a variety of technical options. However, these advancements will only lead to a successful decision support, if the performance analysis and action taking using an OpBI system is consistent to business goals and value creation processes. This requires a conceptual modelling of analytical requirements for OpBI systems in compliance with operational concerns of an organization. We conducted a literature review using the databases of Business Source Complete, IEEE, AIS, ACM, Emerald, and Science Direct to examine scientific publications according to MIS rankings [11]. The reviewed publications do not discuss a conceptual modelling of analytical requirements for a successful application of OpBI. A lack of discussion about conceptual modelling of operational information is evident, yet.

2.2 Related Research Areas

OpBI addresses performance management (PM), BI, and business process management (BPM) [12]. PM structures business strategies and translates them into goals and ratios [13]. Process PM (PPM) monitors business processes using performance indicators [14]. The PPM concept is not limited to a specific IS support, but BPM or BI systems are discussed therein currently [1]. Monitoring business processes has a technical background coming from the BPM perspective [15]. BPM systems log transactions and events for execution tracking and process modelling [16]. The analysis of log data is limited, yet [17]. This extends especially in contexts of sophisticated processes with distributed tasks [18]. Due to an early stage of PM in the area of BPM, an integration of BI and BPM is taken into consideration [3]. From a BI perspective, the analysis of process data has a different focus. Business Process Intelligence supports the design and redesign of processes of an organization [19]. This affects a small range of users making strategic or tactical decisions. In contrast, process-centric BI concerns an integration of BI applications into process executions [20]. This affects the process performance due to accelerations and improvements of a process execution. BI provides analytical information to fulfil process related tasks. This differs from our OpBI understanding by using BI techniques for an analysis and control of business processes. Process-centric BI does not address a consideration of analytical information for an immediate measurement of process performance, an investigation of deviations, or a derivation of control actions.

3 Research Method

We apply an action research method, because this has been used successfully to model business requirements and to align them with IT characteristics [6]. This is similar to our area of discourse by a conceptual modelling of OpBI systems. We extend the methodological knowledge and refer it to a participatory form of action research [21] - researchers and practitioners participate in a research process collectively. The collaboration allows a combination of modelling knowledge with practical experiences about analysing and controlling business processes. Action research supports a solution of immediate performance problems and a consolidation of conceptual knowledge on designing OpBI systems. Participatory action research has been successfully applied, too, in order to ensure that IT implementations result in business benefits [22]. Our intention is quite similar as we want to link the conceptual design of OpBI systems with value-based business requirements. Therefore, we deduce a practicability of participatory action research to deliver a business contribution in consequence of an OpBI-reliant decision-making. In a three-year research period, we performed an iterative and collaborative research process together with four organizations. Assumptions on designing OpBI systems were refined in cycles of diagnosis, action, evaluation, and reflective learning [21]. An approach to link value-based business models and OpBI systems emerged in consequence of our experiences. The approach builds upon the findings and multi-perspective view on requirements engineering of Gordijn and Akkermans [6].

4 Linking Value-Based Business Models and OpBI Systems

Our approach consists of different activities resulting in e^{3}value models [6], a classification of business process requirements according to the work system framework [23] and ADAPT models [24]. Figure 1 classifies the elements of the approach into the perspectives of a value-based requirements engineering [6].

An e^{3}value model describes an exchange of value objects between business actors in a commercial network. Such a network consists for instance of an organization anywhere in a value chain with its potential customers and suppliers. Business actors with an equal value proposition can be grouped to market segments. The value objects to be exchanged in a commercial network are trading items (products, services) in consideration of economic equivalents (money). Value activities model specific performance areas, in which an organization creates or adds value to yield profits. To dig deeper in the particular mechanisms of value activities, we bridge to a consideration of the business process perspective using the work system framework. Both approaches consider an internal and external view on organizations. Table 1 demonstrates the coincidence of e^{3}value and the work systems framework. A work system considers participants carrying out business processes by use of information and technology.

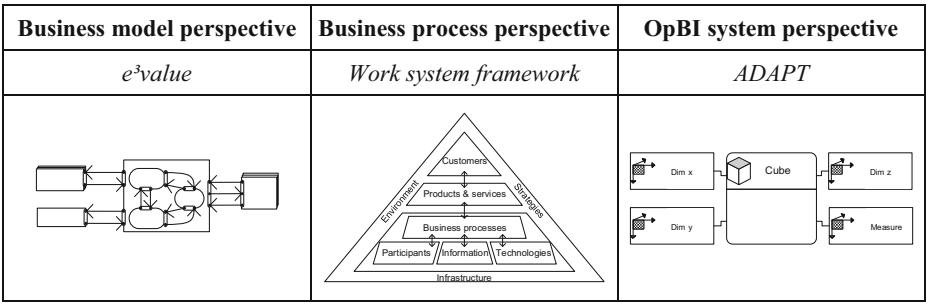


Fig. 1. Elements of our approach

Table 1. Mapping of e^{3}value and work system framework

View on organizations	Elements of e^{3}value	Elements of work system framework
External	The whole e^{3}value model	Strategies
	Customers, external stakeholders, partners, or suppliers modelled as market segments or actors	Environment, customers
	Value exchanges, especially value objects	Products & services
Internal	Concerning organization performing specific value activities	Infrastructure
	Value activities representing areas of performance	Business processes
		Participants
		Information
		Technologies

These core elements of work system performance characterize together with general infrastructure components an insider's view on an organization's business value model. The performance output are products or services, which are the objects of value exchange with customers and the value chain environment. Strategic considerations influence the insider's and outsider's view regarding to work system performance. We use the elements of the work system framework and e^3value to deduce requirements for an analysis and control of value activities from an IT system perspective. Therefore, we use the ADAPT notation to develop logical data models as measurement and structuring instrument for value activity information in operational decision contexts. The work system and e^3value elements are assigned to dimensions and measures of an ADAPT model. The dimensions span a cube consisting of a set of measures having a clear reference to the value objects of the business model. The relationships of dimensions and measures follow the criteria of creating and exchanging values.

5 Action Research Results

We present the results of an action research project that was carried out from August 2012 to February 2014 in Germany in order to develop and apply our linking approach. Four organizations participated in three subsequent cycles of action research. The considered organizations were a machine tool manufacturer, a service provider for IT and communication (ICT) products, a hydraulics engineering company, and an insurance agency. The first cycle refers to activities of interaction, application, and reflection from a business models perspective and results in e^3value models. The outcome of the second research cycle is represented by a work system classification. The third cycle of action research lead to ADAPT models for an OpBI database design. Illustrations of e^3value and ADAPT models are presented only in context of the machine tool manufacturer due to the limited space of the paper.

5.1 Research Cycle 1: Creation of Value-Based Business Models

Machine Tool Manufacturer. The organization modernizes gear hobbing machines. Equipment upgrades happen according to individual customer orders with negotiated budgets, period and quality requirements. The value activities (cf. Fig. 2) include a deployment of new components, such as control units or milling heads. Once the transfer of a customer's machine happens, a dismantling in machine components takes place. Specific and standard parts are cleaned and listed. The employees record geometrical data and take pictures in case of incomplete drawings. Decisions about a rework or a remanufacturing depend on the machine state. Finally, the execution of the re-assembling happens. Disturbance variables are the individuality and the unpredictability of the machines and their states. Different projects and suppliers must be coordinated in consideration of compliance in time and cost conditions.

ICT Service Provider. Logistical services are performed to distribute ICT products from different brands through different channels. The product procurement involves manufacturers or network operators. Devices are customized according to specified

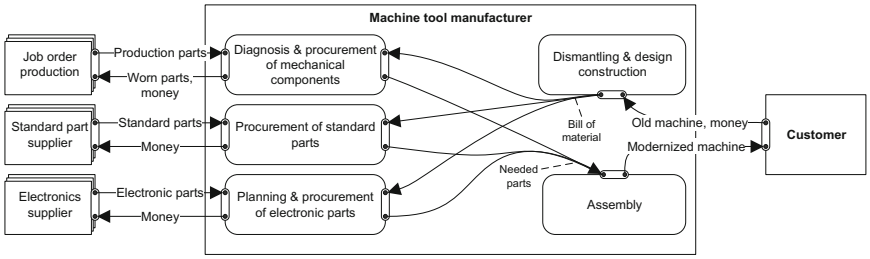


Fig. 2. e^3value model of a machine tool manufacturer

requirements, e.g. exchanges of electronic parts. A recovery resets returned devices to factory settings and performs functional checks. If, necessary, a partner company repairs defect devices. Final products are completed and packaged for shipment. Disturbance variables are fluctuating throughput quantities, a changing staff, heterogeneous products and fast price slumps. Especially velocity and cost efficiency are important control aspects.

Hydraulics Engineering Company. The organization produces hydraulic or pneumatic cylinders and job orders. The customer segment includes industrial trucks, rail vehicles, agricultural machinery, printing presses, or injection moulding machines. The manufacturing concerns activities of metal processing like milling, drilling, turning, welding, or laundry. The assembly of finished products includes functional tests, colouring, and shipment. A quality control records complaints during production and decides for rework, sorting, claim, or scrap. Disturbance variables are order withdrawals, missing materials, troubles of external manufacturers, unavailable labours, or malfunctions of e.g. automatic welders or CNC machining centres. Such disturbances lead to delays of planning cycles. Considering monthly value creation targets should overcome the uncertainties. This means that alternative outputs have to compensate adverse circumstances, if, for example, an order is cancelled.

Insurance Agency. Insurance products are distributed on behalf of an insurance group. The strategic goals of the insurance group concern high premium customer portfolios, optimized trading results, and excellent business processes. The insurance agency has to fulfil the goals by efficient service actions. A planning and scheduling of sales conversations concludes insurance contracts for different products with commercial or individual customers. The agency coordinates, supervises and settles customer claims. Disturbance variables are manifold. Expiring insurance contracts or premature dismissals reduce the number of customers. Failing approaches to agree conversation dates or cancellations counteract attempts to sustain or increase sales revenues. Delays or contradicting information impair the handling of claims due to a missing communication between different contact points, which record claims or requests.

5.2 Research Cycle 2: Classification of Business Process Characteristics

Table 2 classifies the studied organizations into the work system framework. The processes need to be dynamic with a certain variability. The business processes are deterministic and repeatable, while the performance results differ for changing situations. The tasks depend on knowledge and experience of the employees executing, guiding, and instructing operational activities. The information refers to reference inputs, control indicators, resources, products, or stakeholders.

Information technologies mentioned in Table 2 refer to ERP, product data management, warehouse management, or collaborative portal solutions. Important is the availability of data collection techniques. The infrastructure includes a low to medium specialized technical equipment. Human resources are specialists and executive staff

Table 2. Classification of case studies into work system framework

	Insurance agency	Hydraulics engineer	Machine tool manufacturer	ICT service provider
Processes and activities	Consulting, claim settlement, sales conversations	Manufacturing, quality control, assembly	Dismantling, cleaning, rework, assembly	Customization, recovery, shipment
Participants	Senior manager, back office, sales representatives, call centre agents	Engineers, assemblers, operators, supervisors	Project teams with assemblers, engineers, project leader	Shop floor and temporary staff, supervisors and unit manager
Information	Customer records, availability and history, cross selling ratio, claims, expense ratios, premium targets and incomes, contracts	Time data, design drawings, bill of materials, defect reports, article data, consumption rates, target/actual quantities, expense ratios, added value	Time data, design drawings, geometrical data, bill of materials, orders, delivery dates, quality indications, budged limits	Time data, expense ratios, target quantities, delivery dates, article master data, consumption rates, actual quantities, defective products
Technologies	Platform to prepare and manage proposals, policies issues, portfolios and accountings	ERP, Product data management, Machine data acquisition, Time keeping	ERP, Product data management, Time keeping, Project management system,	ERP, Warehouse management system, Machine data acquisition, Time keeping

(*Continued*)

Table 2. (*Continued*)

	Insurance agency	Hydraulics engineer	Machine tool manufacturer	ICT service provider
Infrastructure	Office equipment with interfaces to the insurance group, four employees	Office and production equipment, 100 employees, staff involvement	Office and production equipment, 70 employees, project hierarchies	Office and logistics equipment, 1,500 employees, flat hierarchies
Strategies	Increase of shareholder values, high premium customers	High quality, flexibility and velocity, reliability to customers	Specialization, focus on customer, undercutting of original prices	Diversification of sales, service and repair, high quality at low costs
Environment	Insurance group, financial markets, changing commercial and legal conditions, regional sales area	Supplier relations, high competitive pressure, growing international market	Supplier relations, high competitive pressure, deadline and cost pressure, international market	Supplier and partner relations, international market, varying order situations, fast slumps
Customers	Individual and business clients	Machine building companies	Metal processing companies	Retailers and resellers
Products and services	Insurance products, financial services	Hydraulic cylinder, job orders	Gear hobbing machines	ICT products

organized in problem-oriented communication hierarchies. Customer relations are business-to-business and business-to-customer. The organizations offer specialized products or services in different price segments with a medium to high complexity. They have heterogeneous configurations and consist of sophisticated features. The environment is characterized by competitive pressure and changing conditions in regional and international distribution areas. External factors are the behaviour of suppliers, partners, or associated companies. Strategies of the studied organizations include specialization, diversification, quality excellence, flexibility, velocity, and customer orientation.

5.3 Research Cycle 3: Logical Application Design of OpBI Systems

Machine Tool Manufacturer. The OpBI system supports the budgeting and scheduling of modernization projects. Data gathering happens manually due to the heterogeneity of working activities. A tracking system records the corresponding working times. The database design (cf. Fig. 3) points out expenses for performing the value

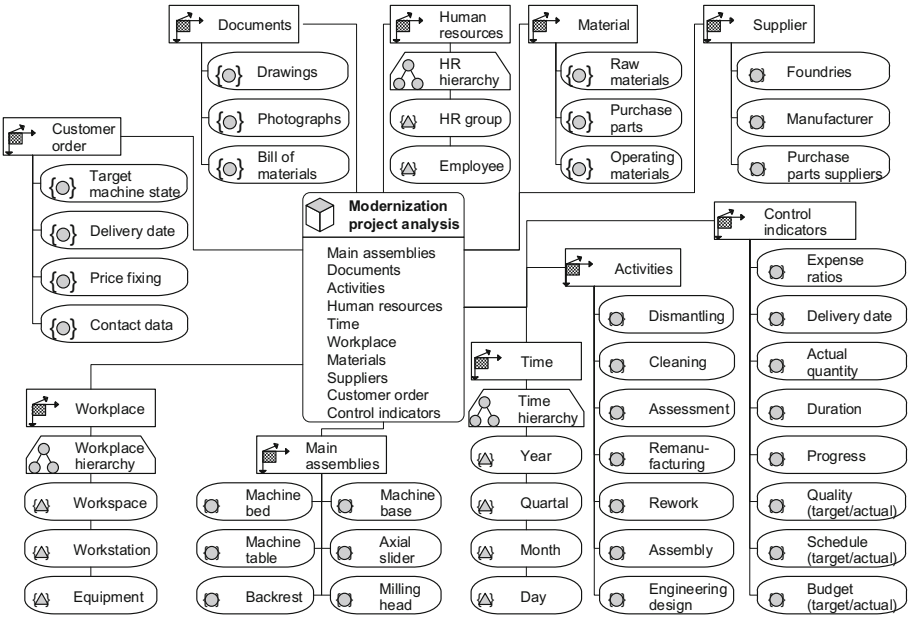


Fig. 3. ADAPT model of a machine tool manufacturer

activities on different levels of detail. Planning and management of project workflows happen simultaneously. Current states of a machine, incurred costs, spent working times, and delivery progress of needed assemblies are demonstrated.

A comparison of actual performances to target indicators enables staff to intervene in case of deviations. The procurement combines supplier information with required rework orders and the quality of finished parts. This rating of providers eases a selection for similar constructed parts. The restored machine features of the individual overhaul projects are comparable so that cost estimations become more confident.

ICT Service Provider. The OpBI system evaluates cost transparency and efficiency to react fast and flexibly on changing order quantities. The affected IT systems are an ERP and a warehouse management system. The data collection occurs with scanners, light barriers, and a machine time tracking. The designed OpBI database provides a basis to derive management actions for an adjustment of order cycles according to product groups. The consequence is a coordination of logistical cost and product-specific price slumps. The data model facilitates a combination of production batches with similar or equal features to improve processing times. Faced by staff changes, performance targets are determined according to human resource groups. These targets depend on product groups and periods of employment. The calculation of product-specific delivery times leads to higher planning reliability as consequence of specifiable agreements for repair services in context of an outsourced repair service.

Hydraulics Engineering Company. The OpBI system determines a value added of manufacturing activities and cost ratios of quality issues. The underlying IT systems

refer to an ERP system with integrated data acquisition. Terminals collect production data using card readers and barcode scanners. A quality assurance tool collects internal quality complaints. The logical designed database supports an incremental accretion measurement of components and products during manufacturing and assembly. This ensures a constant review of value creation targets. Differences will lead to immediate decisions. A consideration of expenses to create specific features improves the employment of resources, materials, and technologies. Constructors get information to determine prices for new products or add-ons during the design phase based on needed product features. The quality assurance derives actions by costs-by-cause principles using the different process perspectives. The logical model enables a calculation of expenses for rework, sorting, or scrap for internal quality complaints.

Insurance Agency. The OpBI system combines information of more than 1,800 customers with allocated service tasks. A platform for proposal preparation, policy issues, portfolio management, and accounting supports semi-standardized information records. Sales representatives or office employees enter this information manually. The OpBI's data model considers reasons for unsuccessful approaches to agree conversations. For example, holidays or shift work lead often to calls at inconvenient customer situations. The scheduling is managed according to reachability of customers, now, and appointments are located in nearby sales regions to reduce travel cost. The data model supports a customer-specific control of claim handling to achieve a well-founded settlement. This depends on extent of loss or damage, underlying insurance contracts, and customer behaviour. The agency monitors deadlines for claim review to accelerate handling times. It is measurable whether a customer has already reported claim information and how far the reports coincide. A comparison of monthly premiums with a number of contracts per customer leads to a prioritization of claims or a consideration of goodwill. This is beneficial to decide about win-back actions in notice management, too.

5.4 Lessons Learned

The action research cycles demonstrate methodological and organizational issues to design OpBI database systems based on value-based business requirements. This delivers insights on measuring and evaluating the performance of business processes in four business scenarios. The conjoint reflection of business models, business processes, and IT systems has proven to be advantageous. Valuable results were achieved in all four organizational settings despite of different situational characteristics. Figure 4 repeats the relation between the perspectives of our approach.

The joined elements of e^3value and work systems are linking an organization's strategy with the maintenance of performant business processes. The association to OpBI is represented at the bottom of Fig. 4. ADAPT models are instruments to collect, elaborate, and analyse data about business processes and build the basis to configure management actions. An important aspect learned from our research is the context-sensitive enrichment of these common descriptive perspectives (cf. Table 3).

The linkage of OpBI with firm-specific business requirements is irreducible complex by observational research methods, because it is necessary to involve situated and practical knowledge resulting from collaboration activities between researchers and

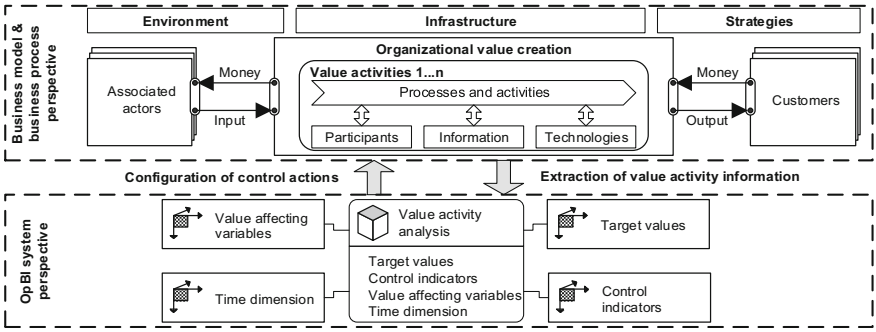


Fig. 4. Relation of business model, business process, and OpBI system perspective

Table 3. Firm-specific and common aspects of our results

Firm-specific aspects	Common aspects
• Business contexts and strategies	• Methodological building blocks
• Business process descriptions	• Action research cycles
• Performance management situations	• Collaboration of research and practice
• OpBI database designs	• Business process orientation
• Management control actions	• Use of operational IT systems

practitioners. Participatory action research enables such a reference to practical contexts. Implications concerning a performance management of business processes depend thereby always on specific organizations. However, our conceptual findings allow a broader consensus on modelling OpBI systems, although they are not object of a rigorous generalization. Especially the work system framework helped us to learn about common aspects like operational IT systems and repeatable business processes.

6 Conclusion

OpBI will support the management of performant business processes, if the analytical concerns are in concurrence to the business requirements of an organization. The paper's contribution enhances a discussion about conceptual aspects of linking OpBI systems design with value-based business requirements. We developed and applied a management approach in coherent action research cycles to provide a conceptual basis for designing OpBI systems from a business perspective.

The paper's arguments shift the discussion about an operational decision-making from technical aspects to a consideration of business strategies. Such a view on information systems is in line with contributions about the impact of IT on business process performance [2]. The novel conceptual approach of value modelling, work system analysis, and analytical design is relevant for application developers and business operation's managers. This supports a definition and evaluation of requirements for an operational decision-making in an organization's business context. The

gained conceptual and practical experience from our action research project refers to four different business scenarios. The collaboration of researchers and practitioners has produced a valid conceptual approach and meaningful outcomes in practical contexts. One learning effect is that a consideration of such collaborative efforts leads to firm-specific implications and to reproducible conceptual insights.

This paper builds its evidence on action research, so that its findings and implications have a qualitative nature. The investigated organizations represent typical scenarios of manufacturing and service provision. This indicates a certain resilience of the action research method and is intercessional for a confident replication logic in additional business scenarios. Upcoming research activities should therefore further consolidate conceptual considerations about the integration of analytical concerns and business value perspectives. This allows taking charge of changing analytical technologies and digital opportunities based on a given business logic or value constellation.

References

1. Blasini, J.: Critical success factors of process performance management systems: results of an empirical research. In: 21st European Conference on Information Systems, Utrecht, vol. 158, pp. 1–12 (2013)
2. Trkman, P.: The critical success factors of business process management. Int. J. Inf. Manage. **30**, 125–134 (2010)
3. Vukšić, V.B., Bach, M.P., Popovič, A.: Supporting performance management with business process management and business intelligence: a case analysis of integration and orchestration. Int. J. Inf. Manage. **33**, 613–619 (2013)
4. Davis, J., Imhoff, C., White, C.: Operational Business Intelligence: The State of the Art. Beye NETWORK Research, Boulder (2009)
5. Hänel, T., Felden, C.: towards a stability of process oriented decision support concepts using the example of operational business intelligence. In: Pre-ICIS BI Congress 3: Driving Innovation Through Big Data Analytics, Orlando (2010)
6. Gordijn, J., Akkermans, H.: Value based requirements engineering: exploring innovative e-commerce ideas. Requirements Eng. J. **8**, 114–134 (2002)
7. McAfee, A., Brynjolfsson, E.: Big Data: The Management Revolution. Harvard Bus. Rev. **90**, 60–66 (2012)
8. Juan-Verdejo, A., Baars, H.: Decision support for partially moving applications to the cloud: the example of business intelligence. In: International Workshop on Hot Topics in Cloud Services, New York, pp. 35–42 (2012)
9. Plattner, H.: A common database approach for OLTP and OLAP using an in-memory column database. In: ACM SIGMOD International Conference on Management of data, Providence, pp. 1–2 (2009)
10. Van der Lans, R.: Data Virtualization for Business Intelligence Systems: Revolutionizing Data Integration for Data Warehouses. Morgan Kaufmann, Waltham (2012)
11. MIS Journal Rankings. http://aisnet.org/?JournalRankings
12. Cunningham, D.: Aligning business intelligence with business processes. What Works **20**, 50–51 (2005)

13. Otley, D.: Performance management: a framework for management control systems research. Manage. Acc. Res. **10**, 363–382 (1999)
14. Kueng, P., Krahn, A.: Building a process performance measurement system: some early experiences. J. Sci. Ind. Res. **58**, 149–159 (1999)
15. Janiesch, C., Matzner, M., Müller, O.: Beyond process monitoring: a proof-of-concept of event-driven business activity management. Bus. Process Manage. J. **18**, 625–643 (2012)
16. van der Aalst, W.M.P.: Process Mining: Discovery. Conformance and Enhancement of Business Processes. Springer Publishing Company, Berlin (2011)
17. Kang, B., Kim, D., Kang, S.H.: Periodic performance prediction for realtime business process monitoring. Ind. Manage. Data Syst. **112**, 4–23 (2012)
18. Cheung, M., Hidders, J.: Round-trip iterative business process modelling between BPA and BPMS tools. Bus. Process Manage. J. **17**, 461–494 (2011)
19. Felden, C., Chamoni, P., Linden, M.: From process execution towards a business process intelligence. In: Abramowicz, W., Tolksdorf, R. (eds.) BIS 2010. LNBIP, vol. 47, pp. 195–206. Springer, Heidelberg (2010)
20. Bucher, T., Gericke, A., Sigg, S.: Process-centric business intelligence. Bus. Process Manage. J. **15**, 408–429 (2009)
21. Baskerville, R.L.: Investigating information systems with action research. Commun. AIS **2**, 1–32 (1999)
22. Breu, K., Peppard, J.: The participatory paradigm for applied information systems research. In: 9th European Conference on Information Systems, Bled, pp. 243–252 (2001)
23. Alter, S.: Work system theory: overview of core concepts, extensions, and challenges for the future. J. Assoc. Inf. Syst. **14**, 72–121 (2013)
24. Getting Started with ADAPT. http://www.symcorp.com/downloads/ADAPT_white_paper.pdf

Operationalizing Data Governance via Multi-level Metadata Management

Stefhan van Helvoirt and Hans Weigand(✉)

Tilburg School of Economics and Management, Tilburg, The Netherlands
svhelvoirt@gmail.com, h.weigand@tilburguniversity.edu

Abstract. Today's rapidly changing and highly regulated business environments demand that organizations are agile in their decision making and data handling. At the same time, transparency in the decision making processes and in how they are adjusted is of critical importance as well. Our research focusses on obtaining transparency by not only documenting but also enforcing data governance policies and their resultant business and data rules by using a multi-level metadata approach. The multi-level approach makes a separation between different concerns: policy formulation, rule specification and enforcement. This separation does not only give more agility but also allows many different implementation architectures. The main types are described and evaluated.

Keywords: Data warehouses · Data governance · Metadata · Business rule enforcement

1 Introduction

The amount of data that is available in the digital universe is growing at an exponential rate and will only continue to grow with the rise of new technologies such as the Internet of Things. Nowadays data is more important than ever before due to the speed of business change. This is emphasized with the rise and use of Master Data Management (MDM) systems in the last decade. MDM adds a new dimension to the data that focusses on establishing integration and interoperability of heterogeneous databases and applications in a business oriented manner [1, 2].

Recent studies have shown that organizations that are capable of effectively utilizing and analyzing their data outperform their competitors [5]. In order to actively use the data that is available both within and outside the organization, the organization must find a way to actively and sufficiently tag the data with metadata [9]. Especially in a new digital world in which organizations are rapidly integrating data from various heterogeneous sources. This need is emphasized with the rise of new data warehouse platforms such as IBM's Data Reservoir. Having a proper data governance program in place is crucial for effectively managing the data that resides in such aggregated environment [3].

What is important in today's highly regulated business environments is not only effective data governance but also that the governance is transparent and auditable. For instance, exporting and importing shippers need to comply with tax regulations and

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 160–172, 2015.
DOI: 10.1007/978-3-319-25013-7_13

customs security controls. It is very hard for companies to prove compliance if the data infrastructure is not well-controlled in a transparent way. Sometimes data governance is mandatory by law as with BASEL BCBS 239, effective from 1/1/2016.

Our research goal is obtaining adaptability and transparency by not only documenting but also enforcing data governance policies and their resultant business and data rules. In this paper, we introduce a multi-level framework and use it to evaluate the current capabilities of IBM's InfoSphere package as used in its Data Reservoir solution, while also providing incentives to further extent the governance capabilities. Additional layers of logic are added to reify governance policies in data movement, applications and databases. A preamble on Data Governance and metadata is provided in Sect. 2, to lay the foundation for our multi-level metadata framework discussed in Sect. 3. Section 4 continues with an overview of various implementation styles to establish a Data Governance environment and Sect. 5 evaluates IBM InfoSphere offerings to establishing operationalized Data Governance.

2 Background

2.1 Data Governance versus Data Management

According to Khatri and Brown [6], based on Weill and Ross, "governance refers to what decisions must be made to ensure effective management and use of IT (decision domains) and who makes the decisions (locus of accountability for decision making). Management involves making and implementing decisions." Data management activities focus on the development and execution of architectures, policies, practices and procedures to enhance and manage the information lifecycle within a specific application and mostly during data entry/creation. Data Governance on the other hand also includes aggregated and integrated data that is made available as a data asset within the organization.

The Data Governance domain consists of three focus areas; people, processes and technology. Many publications on data governance focus primarily on the people and processes aspects of implementing a data governance program. We can use IBM's holistic approach to Big Data governance as an example, which consists out of the following six sets: define business problem, obtain executive sponsorship, align teams, understand data risk and value, implement analytical/operational projects and measure results. Its focus has been primarily on the first four steps. Our focus is mainly on the "implement analytical/operational projects" from a technical perspective as this appears to be research gap. However, all steps in the holistic approach are needed in order to have an efficient and reliable data governance program. Capturing and enforcing business rules, without the proper knowledge of the available data, its value and the interdependencies between data is undoable and undesirable. We do not underestimate the political change that is needed to transform the organization into an information-driven environment. An overall transition needs to be made from thinking and developing individual applications to a unified acceptance and usage of data as the foundation of information and knowledge [8].

2.2 Data Quality and Trust

Governance is more than achieving compliance [7]. Achieving data governance has to do with adopting practices and principles that increase data quality and trust. Having established data quality and trust, the organization can start using their data in a reliable and controlled manner and evaluate its data usage and governing capabilities by implementing appropriate metrics. A valuable data quality standard currently in development is ISO 8000. It focusses on data characteristics and exchange in terms of vocabulary, syntax, semantics, encoding, provenance, accuracy and completeness.

2.3 Metadata as Indispensable Enabler

The importance and utilization of metadata has been increasing rapidly over the last decade as metadata is making its transition from a technical aspect to a business necessity. Metadata is needed for establishing data quality and turning data into understandable information that can be consumed by both business/IT users and software for automation.

Looking at publications on metadata form the past fifteen to twenty years shows that there are various types and classifications of metadata, each with its own specific purpose and granularity. This paper uses the metadata framework as presented by Ron Klein (KPMG) at the 2014 ECCMA conference [7]. This metadata framework consists of three vertical levels and three horizontal levels. The vertical levels are *business*, *technical* and *operational*, while the horizontal levels consist of the categories *descriptive*, *administrative* and *lineage*. Business metadata includes business terms, data owners, stewards, and governance policies and business rules governing the data. Technical metadata is used for tool integration to manage, transform and maintain the data. Examples of technical metadata are database system names, table and column names, code values and derivation rules. Operational metadata contains run-time information e.g. last load, usage statistics and log reports. In short, business metadata has a value and meaning for business oriented users, technical metadata is used primarily by Extract, Transform, Load (ETL) developers while operational metadata is used to provide insights in data usage and rule validation.

3 Multi-level Metadata Management

In the following, we will focus on the policies *behind* the meta-data as such, for instance, policies and access control, policies on quality requirements, or policies on the use of semantic standards, rather than the meta-data tags themselves. Our goal is transforming descriptive data governance policies into implementable rules by using a multi-level metadata approach. The multi-level metadata model consists of four levels as depicted in Fig. 1. Distinguishing these levels leads to maximal adaptability and transparency (cf. [4]).

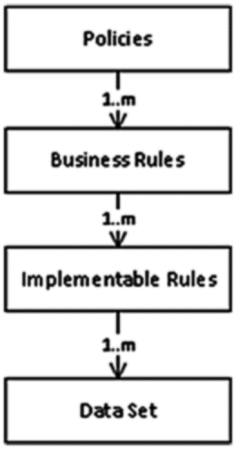


Fig. 1. Multi-level Metadata Management framework

Policies are abstract formulations of business goals, desirable behavior, guidelines and generally accepted practices. Business rules are a formalization of a (partial) aspect of a policy, stating the trigger and follow-up action in a structural natural language consisting of business terminology. Implementable rules are executable objects that contain the logic needed to enforce the business rule. How a business rule is implemented and enforced is entirely dependent on enforcement strategy of that specific rule. As mentioned by Weigand et al. [14], "although business rules are more formal than policies, they are still at the level of business requirements (...), rather than execution. They model "what" is required, rather than "how" it should be implemented" [14]. The same paper emphasizes the need to clarify the enforcement strategy and provides four types of enforcement: preventive, punitive, corrective and adhortative. Here adhortative means that the responsible user is requested to solve the violation when it is detected; the system does not prevent or correct it itself. Punitive means that a sanction is given on violation of the rule whereas corrective means that the system automatically corrects the violation and moves forward to a consistent state, typically by means of compensation. When the enforcement is separated from the business rule specification as such, this allows for great flexibility: the company can switch rather easily from a more loose adhortative approach to a strict preventive approach (or vice versa) depending on the desired compliance levels and operational costs.

Where the enforcement of a business rule takes place is highly dependent on the chosen enforcement strategy and goal of the business rule. For example, a policy stating that all telephone numbers should be formatted according to the applicable standard of the country that it applies to, will have a business rule declaring the use of a given standard for telephone numbers within a given region. This rule could be enforced at the point of data creation using a preventive strategy, or when data is analyzed, transformed and moved to a different location using a corrective strategy.

Business rules are defined as condition action (CA) rules and require a structural transformation to become executable condition action (ECA) rules.

4 Implementation Archetypes

4.1 Business Rule Extraction

For the enforcement of business rules, we start from the generally accepted approach expressed, among others, by Pierre Bonnet in his book on Enterprise Data Governance [2] in which the business knowledge is extracted from the software (hard-coding) and is presented to the business users in an environment that they can (partially) control. "Maintaining knowledge, in particular within complex and evolving organizations that characterize modern companies, cannot survive the trap set out by fixed and stratified hard-coded software, nor informal (textual) documentation, rarely up to date and non-executable". According to Bonnet, a software package must first be able to interact with an MDM system, before it can demonstrate its ability to enforce the relating business rules (BRMS) that will eventually affect the processes (BPM). "First the data, then the rules and finally the processes" [2]. This way, rules are defined per data domain and not based on the software package that uses the data.

4.2 Enforcement Architectures

Isolating data governance rules from the code is one thing, but still leaves many choices on how to enforce the rules. Based on our analysis, we distinguish between a *decentralized*, *centralized* and *leveled* implementation archetype. The archetype that is most applicable to a given situation depends on the available resources and business requirements [15]. For example, an analytical driven environment will have a specific way of enforcing policies as data is collected from various sources and ingested into one or multiple repositories designed and optimized for specific analytical computations (e.g. IBMs Data Reservoir). The enforcement of policies in such an environment could largely occur at the processing of data movement. On the other hand, enforcing governance policies on the actual applications/databases that create/store the data would require a different approach to integrating and enforcing policies. The difference in these three implementation styles as described in this section, is purely in the area of policy *enforcement*. Our base assumption is that policy and asset descriptions are high level and should not be restricted or influenced by the underlying technology and infrastructure. Furthermore, capturing, defining and maintaining the definition and description of policies and assets at domain or organizational level allows for greater consistency, transparency and manageability. However, this integration also has its costs and concerns. One of the concerns is that responsibility for some resource, including data, should not be taken away from the agents owning the data.

We start off by illustrating and defining the decentralized implementation style, displayed in Fig. 2. In this example we have four data storages, each containing the (business) definitions of applicable policies (no pattern) and the resulting implementation code (striped pattern). The decentralized implementation is very common in situations where data governance maturity is low. This implementation style has some benefits and limitations as illustrated in the Table 1.

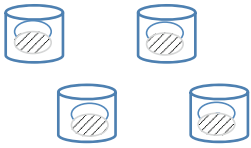


Fig. 2. Decentralized implementation

Table 1. Benefits and limitations of decentralized implementation

Benefits	Limitations
Enforcement of policies as close to the source of data as possible	Siloed knowledge resulting in a lack of reusability and increased risk of inconsistency among separated data storages
Less dependencies and decreased systematic risk	Monitoring compliance and conducting audits is costly and time consuming

Although there are some benefits to mention for the decentralized implementation style, these do not outweigh the limitations. Especially in today's rapidly growing digital ecosystem, in which data is being created by an increased amount of utilities both within and outside the organization. A leveled or centralized implementation style would deliver a more feasible and desirable approach to enforcing data governance, however this requires the presence of a central governance catalog like system for centrally storing and defining data governance policies and assets (cf. [11]).

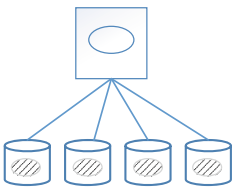


Fig. 3. Leveled implementation

The leveled (Fig. 3) and centralized (Fig. 4) implementation styles make use of a central governance catalog. The leveled implementation style uses a centralized governance catalog repository which is capable of storing the data asset definitions and data governance policies. These asset descriptions describe both the business characteristics of a dataset (business definition using business terminology, owner, steward etc.). Business assets and policies are linked to denote which policies should apply to a specific asset. The implementation and enforcement of these policies is conducted at the System of Record/Reference (SoR). This approach allows for the creation and maintenance of

both asset descriptions and policies at a central level, allowing for greater transparency and consistency. At the same time, the implementation can make use of the tools most efficient for the particular SoR. However, there are also some drawbacks to this implementation style as shown in Table 2. To address the consistency problem, one could imagine an automated update system that pushes any changes in the policy definitions forward to the SoRs. However, when the diverse SoRs use different local enforcement tools, such an update may also require as many compilations as there are different SoRs.

Table 2. Benefits and limitations of leveled implementation

Benefits	Limitations
Consistency in asset and policy definitions	Gap between the definition of a policy and the actual implementation which could result in misinterpretation and incorrect enforcement
Increased transparency in the available data and the rules that shape the data and its use throughout the data lifecycle	Lack of consistency in the enforcement of policies due to high diversity of SoR sources

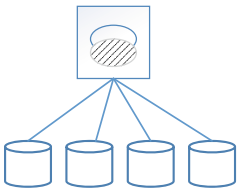


Fig. 4. Centralized implementation

Lastly, we have the centralized implementation style as displayed in Fig. 4. A centralized implementation requires the presence of a governance catalog repository with enhanced and additional capabilities. We can distinguish at least three variants of centralized systems capable of establishing and enforcing governance policies on data assets. These variants differ in how users get access to the data: distributed or intermediated. The first being an environment in which end users direct their requests to the various source applications or databases that in turn call a *service* running on the governance catalog system to evaluate, conduct and if needed enforce a policy. In the second environment, the governance catalog functions as an *intermediary* that ingest data from various sources and uses ETL practices to conduct and enforce governance policies before providing the data to the requesting end user. Lastly is a *hybrid* environment of having both the capabilities of data movement/integration (second environment) and service-like enforcement (first environment).

The hybrid environment contains the highest level of complexity to implement as it requires functionality for two entirely different environments. The first service-based environment excels in a landscape in which policies are defined for enforcing rules designed for the data creation phase (rather than retrieval phase), while the second "intermediary" environment excels in a more analytical landscape in which data needs to be collected, aggregated and delivered to an end user or analytical application (e.g. SPSS). In the first environment we described, the governance catalogs functions primarily as a rule engine. In the second environment the governance catalog functions like a true catalog that controls the data flows from source systems to end users based on the governance policies that are defined. Both environments are needed to establish a holistic data governance solution. Table 3 summarizes the benefits and limitations of a centralized implementation strategy.

Table 3. - Benefits and limitations of centralized implementation

Benefits	Limitations
Optimal consistency in asset and policy definitions and enforcement	May lead to higher network load, possibly lower enforcement efficiency, Single Point of Failure
Allows full integration of policies	Requires high level of integration (organizational and technical)

Technological advances and the use of Enterprise Application Integration (EAI) and Service Oriented Architecture (SOA) to develop new applications and services help in establishing this service-oriented environment for discovering and analyzing of data. EAI allows for the extraction of business policies and rules from the applications, creating increased flexibility and agility. SOA is a framework to "address the requirements of loosely coupled standards-based and protocol-independent distributed computing, mapping enterprise information systems appropriately to the overall business process flow" [10]. Technological advances include, amongst others, new ways of processing data (e.g. NoSQL, in-memory, Hadoop), a decrease in storage costs and increase in memory and computing power to perform the needed operations and a move to semantic systems. SOA can be enhanced with semantic technologies, for instance, to improve service identification [13].

4.3 Catalogue Virtualization

Although a centralized governance catalogue has important management advantages, the drawback is that business users – in particular, the managers responsible for the data –are set on a distance. This can be remedied by virtualizing the catalogue. This means that the various data policies are stored in a distributed way, under the control of the business user. These business users are at various levels: company-wide standards are maintained at corporate level, other policies at division of department level. In the simplest form of virtualization, these distributed data

policies are just synchronized regularly with the central governance catalog. Alternatively, there is only a virtual central catalog, the combination of all distributed policies. In both cases, we assume that policy owners receive feedback (dashboard) on the actual policy compliance.

A critical issue in such a solution is the consistency of the policies. Policies may be conflicting. For instance, a corporate policy may be that all management reports are readable for the internal audit group, whereas a manager may want to restrict access to members of his own department only. One business user may want to express weights in kg and another one in pounds. In the context of this paper, we just mention a few alternative solutions which roughly correspond to the general rule enforcement strategies that we mentioned in Sect. 3. One is to accept inconsistencies as a fact of life and include meta-rules for solving them. A meta-rule can be based on the company hierarchy where corporate policies overrule local ones. This corresponds to a corrective approach because it effectively makes changes in the policies – not in their formulation, but in their application. Alternatively, we can take an adhortative approach that accepts inconsistencies but stimulates policy owners to avoid them at specification time. Closely related, a lazy evaluation (corresponding to a detective approach) can be used that detects conflicts when they actually occur and reports them back to the policy owners. This can be a pragmatic approach in situations of relative low governance where the probabilities of actual conflicts are low. Finally, the most rigid approach is to prevent any inconsistency by using a consistency checker before any policy is deployed. This is a challenge in a distributed environment, although in principle, such a checker is not different from the checkers in a centralized catalog. Last but not least, it is not necessary to choose only one approach. For instance, the company may use a preventive approach for all data standard policies and a detective approach for data access policies.

Once a virtual solution is in place, a next step can be to relax the centralization of the governance catalog. In large companies, a completely centralized approach is not realistic. Some distribution in “regions” or “zones” is unavoidable. In such a situation, a business user may be connected to one region, but also with more regions. Locally, he can manage his policies for both. Data traffic between regions is based on agreements that appear as policies in each of the regions involved.

The virtual solution described in this section can be combined smoothly with a strict distinction between “policy” and “rule” level, as sketched in Sect. 3. This means that the business users publish policies in a user-friendly policy language that is translated to formal business rules on the central catalog (physical or virtual).

5 State-of-the-Art Governance Solutions: IBM InfoSphere

In this section, we analyze in depth one commercially available solution in data governance, IBM InfoSphere. Since this is considered state-of-the-art technology, it can be seen as representative. Our goal for this study was not to compare it with other products, but to see to what extent a multi-level governance model is or can be implemented with this solution. Our analysis is based on the system documentation, expert interviews, and user experience.

5.1 Description

Our evaluation of the IBM InfoSphere suites capability to define and enforce governance policies focuses on IBM InfoSphere Information Governance Catalog (formerly known as InfoSphere Business Information Exchange) and IBM InfoSphere DataStage. Additional tools such as IBM InfoSphere Information Analyzer, IBM InfoSphere Optim and IBM InfoSphere Guardium are used to illustrate specific enforcement examples. Information Governance Catalog is designed to contain both the business glossary (terminology) as well as a list of all available information assets (e.g. dataset, table, policies, and rules) and a variety of additional metadata to describe and define the asset. An information asset is defined as "a body of information, defined and managed as a single unit, so that it can be understood, shared, protected and exploited effectively. Information assets have recognizable and manageable value, risk, content and lifecycles" [12]. Information Governance Catalog allows for the creation of a hierarchical structure to define the relations between policies and rules. These rules can be assigned to a business term. The business term defines and references to the actual source of the authoritative data.

Enforcing data governance policies focusses primarily on achieving a compliance layer. The compliance layer consists of four areas; Policy Administration, Policy Implementation, Policy Enforcement and Policy Monitoring. A *policy* is a (natural language) description of business intent for a class of assets to adhere to a

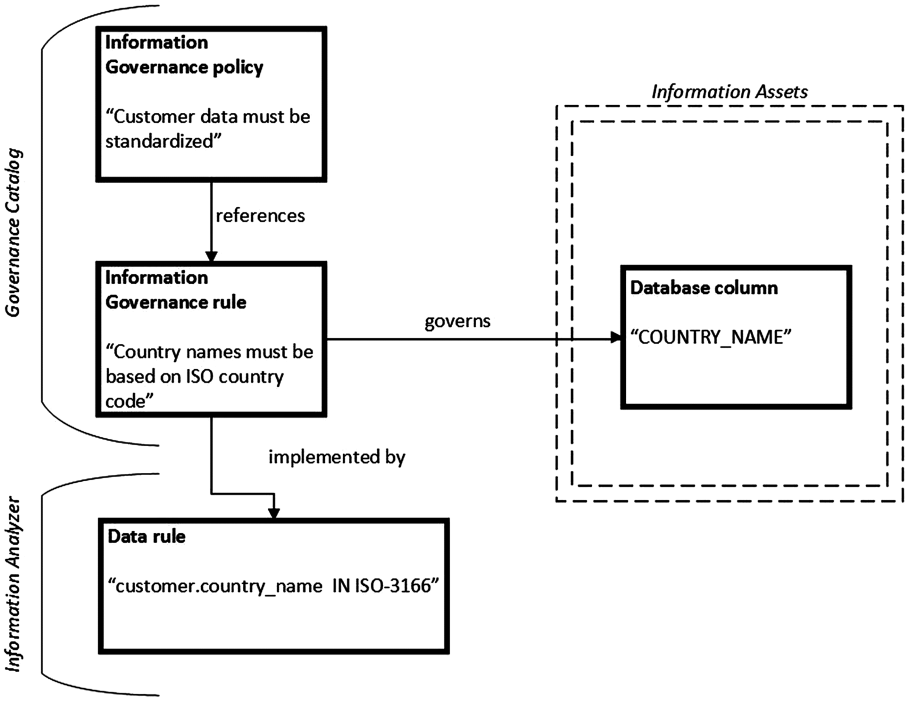


Fig. 5. Infosphere information governance

certain behavior. A *rule* defines how a policy will be implemented, it contains the policy response. Additionally to policies and rules are the *control* and *enforcement points*. A control point is a collection point for evidence that a policy is being complied with. An enforcement point is where a policy implementation (rule implementation or process) is executing. Enforcement points can be seen as hard enforcement measures that guarantee preventive compliance. Control points provide the soft enforcement which is used in the remediation that occurs after the fact. Soft enforcements are used in policies defined at a more abstract level that cannot be hard enforced.

Figure 5 illustrates how information governance policies, governance rules, data rules and the information assets are related. There is some correspondence to the multi-level meta-data framework described above. One difference is that the information assets are related to the information governance rules (business rules), rather than implementable data rules, but if there is 1–1 relationship between data rule and information governance rule, the two representations are equivalent. However, there is no formal representation for the information governance policies and rules, and hence it is not possible to check the consistency of the rules or adapt them automatically.

5.2 Evaluation

Does the IBM InfoSphere suite support multi-level metadata to enable governance, and to what extent? To answer this question, we have looked at the capabilities to validate and enforce rules in the process of moving data (ETL) by using primarily IBM InfoSphere DataStage. Within DataStage we have the capability to create jobs for performing various ETL activities. These jobs can be assigned to a rule in the Governance Catalog as the implementable artifact. Executing a DataStage job results in the creation of operational metadata which is used to establish lineage and provide metadata to the governance dashboard. The operational metadata is mapped to IBM's private proprietary metadata model called XMeta. Besides generating data lineage the capabilities include measuring data quality and values using IBM InfoSphere Information Analyzer. Having insights into the quality and usage of data creates a tremendous increase in transparency for both business and technical users.

The Information Governance Catalog should be used as the central storage point for all the metadata that is needed for providing sufficient insights in definitions, usage, accountability and compliance. However, the current capabilities of enforcing rules and measuring their results requires a lot of technical expertise (ETL development etc.), which is undesirable in an environment that should be business driven. A more formal (semantic) approach to defining the rules should empower the business users with more capabilities to governing "their" data. Policy and rule administration are currently defined in free-text format, which could result in misinterpretations during the implementation and enforcement phase, and creates an opaque environment. A BRMS that utilizes the capabilities of defining rules in a natural structured language reduces opaque and misinterpretation, resulting in a more transparent environment. Answering the research question: IBM InfoSphere suite supports multi-level metadata to enable governance, but there is still a lot of room to enhance these

capabilities to increase transparency and formalization. At the moment, it supports typically a leveled approach, not full centralization.

6 Conclusion

In order to operationalize data governance, the implementing organization needs to have the resources and capabilities in place to define and enforce data governance policies and rules. Using a multi-level metadata framework we created an insightful segregation between defining the policy and rule specifications and the resultant implementation of rules and jobs. With this segregation in place, and the capability of empowering qualified business users to define governance specifications, allows for better adaptability and transparency of data governance. As this paper presented, there are various ways of implementing a multi-level metadata framework. Having the capability to enforce policies and rules both in a centralized and decentralized manner, allows for the most flexibility. However, specific software might need to be purchased to establish an environment for this in the form of a governance catalog. IBM InfoSphere suite provides most of the capabilities needed to start operationalizing a multi-level data governance program, but formalization of policies and rules is needed in order to get to a higher level, in particular, to one supporting self-adaptation.

References

1. Baca, M., Gill, T., Gilliland, A.J., Whalen, M., Woodley, M.S.: Introduction to Metadata, Revised edn. Getty Publications, Los Angeles (2008)
2. Bonnet, P.: Enterprise Data Governance: Reference and Master Data Management Semantic Modeling. Wiley, New York (2013)
3. Cheong, L., Chang, V.: The need for data governance: a case study. In: Toowoomba: 18th Australasian Conference on Information System (2007)
4. Gong, Y., Janssen, M.: From policy implementation to business process management: principles for creating flexibility and agility. Gov. Inf. Q. **29**, S61–S71 (2012)
5. IBM Center for Applied Insights. Outperforming in a data-rich hyper-connected world. New Orchard Road: IBM Corporation (2012)
6. Khatri, V., Brown, C.V.: Designing data governance. Commun. ACM **53**(1), 148–152 (2010)
7. Klein, R.: Metadata is 'not' a technical term anymore: frame to work. In: 2014 International Data Quality Summit, ECCMA 2014: KPMG, pp. 1–18 (2014)
8. Marco, D.: Practical steps for overcoming political challenges in data governance. In: IDQSummit 2014 ECCMA 2014: EWSolutions, pp. 1–37 (2014)
9. NISO. Understanding Metadata. National Information Standards Organization (2004)
10. Papazoglou, M., van den Heuvel, W.-J.: Service oriented architectures: approaches, technologies and research issues. VLDB J. **16**(3), 389–415 (2007)
11. Singh, G., Bharathi, S., Chervenak, A., Deelman, E., Kesselman, C., Manohar, M., .Pearlman, L.: A metadata catalog service for data intensive applications. In: Supercomputing ACM/IEEE Conference (2003)
12. The National Archives. The Role of the Information Asset Owner: a Practical Guide. National Archives (2010)

13. Vitvar, T., Peristeras, V., Tarabanis, K.: Semantic Technologies for E-Government: an Overview. Springer, Berlin (2010)
14. Weigand, H., van den Heuvel, W., Hiel, M.: Business policy compliance in service-oriented systems. Inf. Syst. **36**(4), 791–807 (2011)
15. Wende, K., Otto, B.: A Contingency Approach to Data Governance. In: MIT Information Quality: ICIQ (2007)

A MapReduce Based Distributed Framework for Similarity Search in Healthcare Big Data Environment

Hiren K.D. Sarma[1], Yogesh K. Dwivedi[2(✉)], Nripendra P. Rana[2], and Emma L. Slade[2]

[1] Department of Information Technology, Sikkim Manipal Institute of Technology, Rangpo Sikkim, India
hirenkdsarma@gmail.com
[2] School of Management, Swansea University, Swansea, SA2 8PP, UK
{y.k.dwivedi,n.p.rana}@swansea.ac.uk, emmalslade@hotmail.co.uk

Abstract. Similarity search in the big data environment is a challenging task. Patient Similarity search (PaSi) is an important issue in healthcare network and data. The results of PaSi search may be highly useful for drawing different conclusions and decisions to improve healthcare systems. Such findings can also be useful for choosing the treatment paths for new patients. In this paper, we propose a MapReduce based framework as a solution to the PaSi problem in the context of a healthcare network imagined to be implemented considering the healthcare centers of India. It is assumed that such a healthcare network will be implemented in future over the Government of India cloud known as GI cloud or 'MeghRaj'. The paper also discusses the associated implementation challenges of the proposed framework and the query handling approach for the proposed framework to solve the PaSi problem is stated. Finally, the paper outlines the future scope of the work.

Keywords: Big data · MapReduce · Similarity search · Patient similarity (PaSi) · Cloud · Framework

1 Introduction

In today's world, the volume of digital data generated by different information and communication technology (ICT) related applications, and other applications in the domain of meteorology, scientific instruments, healthcare or medical networks, etc., is enormous [1]. Such huge volume of data, which is generated largely over networks, is not practical to handle through classical database management approaches [3]. For these systems, capturing, storing, processing and retrieval of appropriate data in a timely manner are some extremely important issues. Centralized solutions to these problems are not suitable and distributed solutions have their own problems [2]. Some problems of distributed processing include network bottlenecks, requirements of global information locally, extra communication overheads, etc. Such applications have given birth to the concept of big data. Big data has attracted the attention of researchers and data scientists in recent times. Novel solution approaches are required to handle big data related issues [1].

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 173–182, 2015.
DOI: 10.1007/978-3-319-25013-7_14

Healthcare network is one example, which generates a huge volume of data every day. One of the processing issues connected to healthcare data is to find out Patient Similarity (PaSi). It is defined as the rate of similarity between two or more patients in terms of their symptoms, treatment procedures, personal information, etc. [1]. A typical PaSi solution will find out those patients who have the greatest amount of information in common. Then the treatment paths followed for such patients can be adapted for new patients. These data related to patients are stored in different databases of patient information systems maintained across healthcare networks. There could be several issues in solving PaSi. The data format used for different patients could be different. This is due to the lack of predefined record structure applicable to all patients. The volume of such data to be processed will be colossal. Moreover, some data related to patients can be uncertain and the data will be generated at a very high rate across the healthcare network. As a result, we need big data solutions to address the PaSi problem of healthcare networks. Hence, there is a need to think of some distributed and scalable solution approaches in order to address this problem. MapReduce is a tool that can be used to develop distributed and scalable solutions against big data problems [2, 3]. MapReduce has also been used to solve some healthcare problems [1].

'MeghRaj' is a cloud computing environment developed by the Government of India [4]. There is scope to implement a healthcare network connecting different health centers or hospitals spread across the country over this cloud. If such a system is implemented, it is going to generate big data. Thus, we need different big data solutions to handle different issues related to data processing and storage of these data.

In this paper, we consider a healthcare network that can be implemented over the 'MeghRaj' cloud and address the issue of finding PaSi. We propose a framework, which is based on MapReduce, to address the PaSi problem. We assume that the patient information will be stored in an unstructured manner. Even in the same machine or data source, two different patients' information can be differently structured. This framework is a proposal and its performance evaluation through simulation is undertaken by us.

The rest of the paper is organized as follows. Section 2 presents the background on big data followed by Sect. 3 in which related works are mentioned and the problem undertaken here is stated formally. In Sect. 4 the proposed framework is discussed and Sect. 5 describes the implementation challenges present in the proposed framework. Finally, Sect. 6 concludes the paper with an outline of the future scope of this work.

2 Background

MapReduce is a programming model and an associated implementation for processing and generating large datasets [3]. It is possible to handle big data through MapReduce, programmers find the system easy to use, and this parallel data processing tool has been made popular by Google. It is a scalable and fault-tolerant data processing tool that makes it possible to process a massive volume of data in parallel with the association of many low end computing systems [2]. Users specify the computation task at hand in terms of a map and a reduce function, and then the underlying runtime system processes the given task by distributing the computation tasks across large scale clusters of

computation nodes. This tool can handle machine (i.e. computation node) failures and can also make efficient use of network and disks by appropriate scheduling mechanisms. A decomposable algorithm, partitionable data, and sufficient small data partitions are required for effective use of MapReduce [6]. There are some enhancements to MapReduce. For example, in the work [5], classic MapReduce was optimized to decrease the data transformation load. A shared area for information was considered in this approach. Such an approach is suitable for solving problems like k-nn and top k queries. In [8], a method was developed to handle workloads in hierarchical MapReduce architecture. Haloop proposed in [7] is another type of MapReduce structure suitable for handling iterative problems. iMapreduce proposed in [10] also supports iterative processes. The work presented in [12] is aimed at reducing the amount of data transferred in the MapReduce network. Here, MPI (Message Passing Interface) was used for message passing in a MapReduce structure. The work presented in [11], replaces Hadoop File System (HDFS) with a concurrency optimized data storage layer. This layer is based on the BlobSeer data management service. It is essential to estimate the input/output (I/O) behavior of MapReduce applications and the work presented in [9] is a model that can be used to estimate I/O behavior of MapReduce applications.

In this section, we also establish the relationship of healthcare data with big data. If we look at the networked environment considering the hospitals across a country like India, then we visualize that patients' data will be generated at an exponential rate. These data will have different formats and standards. In healthcare networks, various data related to patients' health, diseases and recovery processes could be made available. Such data will be of great help for the treatment of other patients. Of course, for this to happen, there is a need of processing these data from different perspectives.

Big data is characterized by four 'Vs' namely volume, variety, velocity, and veracity [1]. Data generated through healthcare networks exhibit all the above four characteristics. In such systems huge volume of data is generated in various formats with a high velocity. Moreover, for many patients we get uncertain data in the data generated by healthcare networks and this fact leads to veracity of healthcare data. Thus all four Vs of big data are present in healthcare data. Therefore, big data solutions are required to solve different data processing problems of healthcare data.

Looking at the high volume of data in healthcare networks, big data solutions are necessary for data analysis [1]. According to [13], processing costs can be reduced by using big data analytics in healthcare. In the work presented in [14], problem like selection of appropriate treatment paths is addressed and solution for improvement of healthcare systems has been proposed. A scalable knowledge discovery platform for healthcare big data is proposed in [15].

3 Related Work and Problem Statement

Finding Patient Similarity (PaSi) is the major task considered in this paper. We consider a very specific healthcare network system, which is yet to be built but must be a reality in the near future. The system under consideration is the healthcare network of India. Different healthcare units i.e., hospitals spread across the country will be connected in a hierarchical

manner. One has to find out PaSi of two or more patients in terms of their symptoms, treatments, personal information, etc. The objective in PaSi is to identify those patients who have the greatest amount of information in common. Using the result of PaSi, new patients can be treated by following treatment processes adapted for those previous patients.

PaSi solutions can be found out by either of the two approaches as mentioned in [1]. First is the use of machine learning and data mining algorithms and the second being information retrieval by simple search or by entity-relationship graphs. A brief survey related to these two techniques can be found in [1].

In [1], a MapReduce based method for finding PaSi solution is proposed. The method is scalable and distributed - named as ScaDiPasi - takes small execution time and is implementable over big data related to healthcare networks. The experimental results reported in the paper show that the ScaDiPasi would be able to produce PaSi solutions over big data of healthcare networks.

3.1 Cloud of Government of India

With an aim of exploiting the benefits of cloud computing, the Government of India has initiated a very ambitious GI cloud project. This cloud has been named as 'MeghRaj' [4]. As mentioned in [4], the objectives of GI cloud are: optimum utilization of infrastructure; speeding up the development and deployment of e-governance applications; easy replication of successful applications across different states of the country to avoid duplication of effort and cost in development of similar applications; and, making the certified applications following common standards available in one place.

The GI cloud consists of multiple national and state clouds. A detailed discussion on the architecture of GI cloud and projects to be implemented under GI cloud can be found in [4].

This GI cloud infrastructure can definitely be used for implementing a healthcare network across the country. Such a healthcare network will generate healthcare big data. Any stakeholder of the healthcare network can share the benefits of processing such big data with different objectives within less time. Therefore, the advantages and benefits of such a system can be significant.

3.2 Problem Statement

Finding Patient Similarity (PaSi) is an important problem. As already discussed, there are different approaches for finding PaSi solutions. Interestingly, there exists no single solution applicable to all kinds of problems. Although there are several PaSi solutions already proposed for different situations, we believe, those will not be directly applicable to the big data to be generated by the cloud based healthcare network system of India that is being considered in the current research. Therefore, we need a novel solution to solve the PaSi problem of the healthcare network over GI cloud. Hence, the problem statement of this paper is:

> To design a big data based framework for addressing the Patient Similarity (PaSi) problem considering the future healthcare network that can be built over the GI cloud.

4 Proposed Framework

In this section, we provide a framework for similarity search related to the healthcare big data environment. The proposed framework is based on MapReduce [3]. We consider the cloud environment 'MeghRaj' [4]. It is assumed that a healthcare network considering all health centers of India spread across the country will be implemented over 'MeghRaj'. The structure of the healthcare network will be hierarchical as shown in Fig. 1. Different layers in this hierarchy are the hospitals at different levels such as (i) at panchayat level, (ii) at block level, (iii) at sub-division level, (iv) at district level, (v) at state level, (vi) at region level, and (vii) at national level. Although the type of hospitals can be categorized as elementary health center, public health center, medical college and hospital, private hospital, etc., at this stage we do not discriminate the hospitals based on type. We assume that all hospitals are similar with respect to generation of patients' data. We mainly need to work on the patient data irrespective of the facilities and infrastructure of the hospitals, which are generating such data.

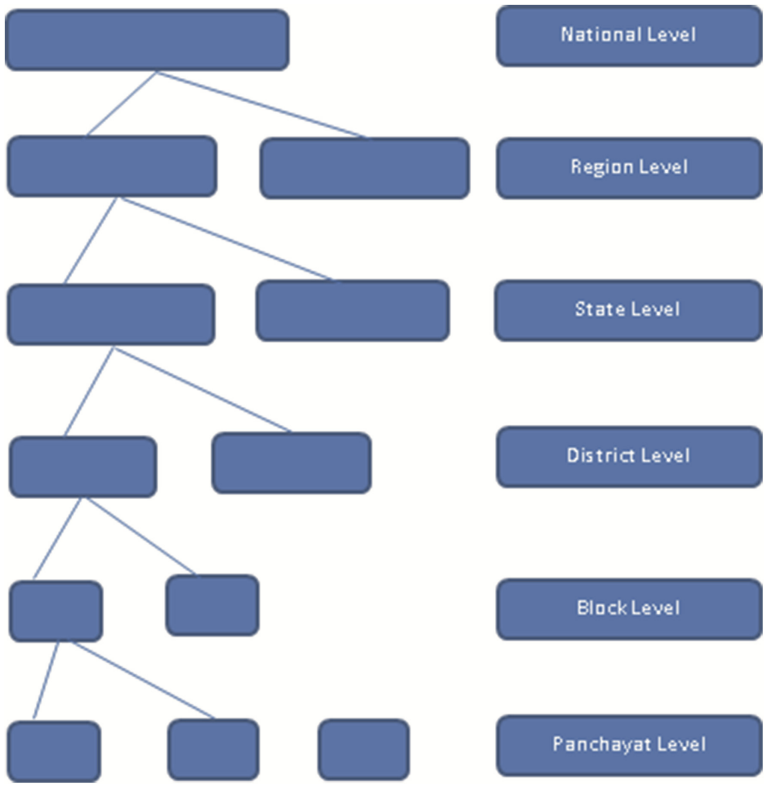


Fig. 1. Hierarchical organization of health centers

The data generated by such a healthcare network system will be stored in the cloud 'MeghRaj'. It is assumed that the healthcare network will be implemented over the Internet. As an end result, a legitimate user of the healthcare network should be able to throw a query related to similarity search problem to the healthcare network and should receive the result back from the network as quickly as possible. This scheme is shown in Fig. 2.

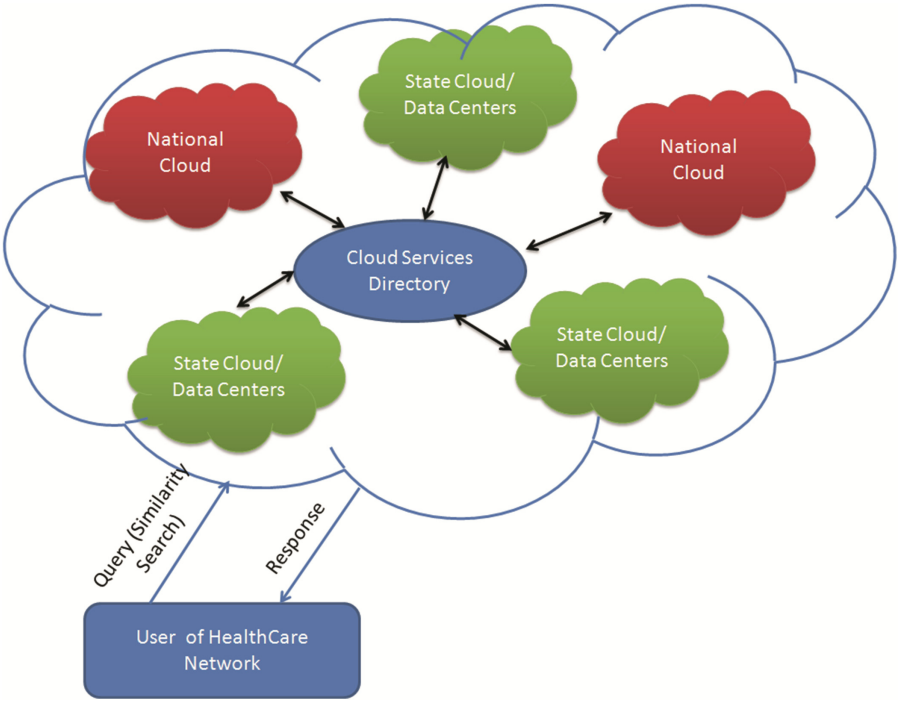


Fig. 2. Query sending to healthcare network to be implemented over MeghRaj cloud

4.1 How MapReduce Will Work

In order to optimize the solution process of similarity search problem, we propose a load-balancing module, which keeps track of distributed load among the computing nodes of the network. Moreover, this module tries to balance the computing load among the nodes.

As we assume that the data sources can be of heterogeneous nature, we propose to have a data format middleware, which brings different data formats to a homogenous common format before any kind of processing task takes place.

Bottlenecks due to maximum message exchanges in distributed processing are an issue, which is also unavoidable. We propose a bottleneck assessment and control module that takes care of the possibilities of bottleneck occurrence. In the presence of

a bottleneck in certain nodes, this module will divert the necessary workload to some other nodes so that the bottleneck can be controlled temporarily.

Data aggregation module will aggregate different data against different queries into some aggregated state. In an aggregated state the volume of data will be reduced significantly and intermediate code to represent data will be generated. This aggregated data in encoded form with reduced volume will be moving across the network reducing amount of data traffic in the network.

All these four modules mentioned above i.e., load balancing module, data format middleware, bottleneck assessment and control module, and data aggregation module, will be working outside the MapReduce environment. The output of these modules will be integrated with MapReduce for optimal performance of the entire system. This framework is depicted in Fig. 3.

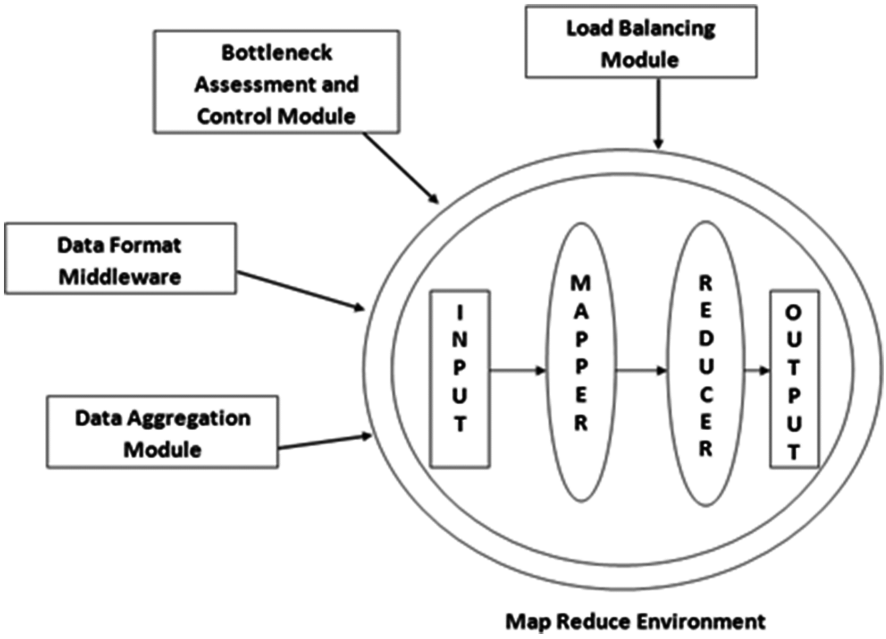


Fig. 3. Proposed MapReduce based framework for similarity search in healthcare big data

5 Implementation Challenges

The major challenges in total implementation of the proposed framework are as follows:

Challenge Set 1. Appropriate technique for load estimation and appropriate algorithm for load balancing are to be designed and associated theoretical complexity analysis is to be carried out. Proper task scheduling algorithm for load balancing is to be designed and analyzed.

Challenge Set 2. Data format middleware is to be designed, which will be highly specific to the structures of different databases present in the healthcare network.
Challenge Set 3. Proper algorithm for bottleneck assessment is to be designed. Moreover, bottleneck has to be controlled and this may lead to migration of processes or computation tasks from one node to another lightly loaded node. Thus there is a necessity to design proper process migration algorithm.
Challenge Set 4. Data aggregation algorithm considering the patient databases is to be designed and analyzed for its performance. Proper encoding mechanism has to be designed for it.

5.1 Solution for Similarity Search

The Patient Similarity (PaSi) search is the similarity search problem considered here. A user throws a query, and then this query is translated into an appropriate uniform format through an intermediate query building process. It works on the MapReduce framework. We propose to have three phases of query processing to solve the PaSi problem. In each phase, Mapper and Reducer functions are to be implemented along with a Ranker function. The Ranker function at each phase evaluates the similarity level of the output of each phase with the input query. This is proposed to be a dynamic decision regarding forwarding of the output of the previous phase to the next phase for further processing in search of more similar results. A threshold level of similarity can be set as per the wish of the user with respect to his/her query. The second phase will continue to be executed until the similarity equates or exceeds the threshold level set by the user. This evaluation is carried out by the Ranker function to be implemented in each phase and a decision is made dynamically regarding the forwarding of the processing task to the next phase. This scheme is shown diagrammatically in Fig. 4.

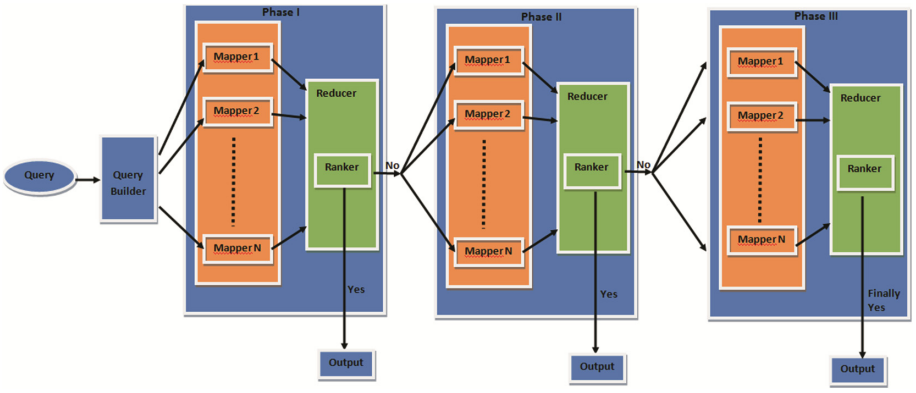


Fig. 4. Proposed query processing model based on MapReduce

6 Conclusion

In this work we address the problem of similarity search in a healthcare network using 'big data'. We focus on the Patient Similarity search popularly known as PaSi, and propose a framework for addressing the PaSi problem in healthcare data. Implementation issues are discussed thoroughly and a MapReduce based model of query handling is also proposed. The proposed framework and the query handling model are designed considering the Government of India cloud also known as 'MeghRaj'. It is assumed that a healthcare network will be implemented considering all the hospitals spread across India and will be deployed over 'MeghRaj' cloud. As for future scope of this work, it is noteworthy that various algorithms can be designed to address the implementation challenges outlined in Sect. 5. Moreover, the query handling model can be implemented over MapReduce framework considering some suitable patients database and various performance parameters, like execution time and accuracy, can be measured and analyzed.

References

1. Barkhordari, M., Niamanesh, M.: ScaDiPaSi: an effective scalable and distributable MapReduce-based method to find patient similarity on huge healthcare networks. Big Data Res. **2**(1), 19–27 (2015)
2. Lee, K.H., Lee, Y.J., Choi, H., Chung, Y.D., Moon, B.: Parallel data processing with MapReduce: a survey. AcM sIGMoD Rec. **40**(4), 11–20 (2012)
3. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
4. GI Cloud Initiative (2015). http://deity.gov.in/content/gi-cloud-initiative-meghraj
5. Ding, L., Xin, J., Wang, G., Huang, S.: ComMapReduce: an improvement of MapReduce with lightweight communication mechanisms. In: Lee, S.-G., Peng, Z., Zhou, X., Moon, Y.-S., Unland, R., Yoo, J. (eds.) DASFAA 2012, Part II. LNCS, vol. 7239, pp. 150–168. Springer, Heidelberg (2012)
6. Highland, F., Stephenson, J.: Fitting the problem to the paradigm: algorithm characteristics required for effective use of MapReduce. Procedia Comput. Sci. **12**, 212–217 (2012)
7. Bu, Y., Howe, B., Balazinska, M., Ernst, M.D.: HaLoop: efficient iterative data processing on large clusters. Proc. VLDB Endowment **3**(1–2), 285–296 (2010)
8. Martha, V.S., Zhao, W., Xu, X.: h-MapReduce: a framework for workload balancing in MapReduce. In: 27th International Conference on IEEE Advanced Information Networking and Applications (AINA), pp. 637–644 (2013)
9. Groot, S.: Modeling I/O interference in data intensive Map-Reduce applications. In: 12th International Symposium on IEEE/IPSJ Applications and the Internet (SAINT), pp. 206–209 (2012)
10. Zhang, Y., Gao, Q., Gao, L., Wang, C.: Imapreduce: a distributed computing framework for iterative computation. J. Grid Comput. **10**(1), 47–68 (2012)
11. Nicolae, B., Moise, D., Antoniu, G., Bougé, L., Dorier, M.: BlobSeer: bringing high throughput under heavy concurrency to hadoop Map-Reduce applications. In: 2010 IEEE International Symposium on Parallel & Distributed Processing (IPDPS), pp. 1–11 (2010)
12. Mohamed, H., Marchand-Maillet, S.: MRO-MPI: MapReduce overlapping using MPI and an optimized data exchange policy. Parallel Comput. **39**(12), 851–866 (2013)

13. Srinivasan, U., Arunasalam, B.: Leveraging big data analytics to reduce healthcare costs. IT Prof. **15**(6), 21–28 (2013)
14. Jee, K., Kim, G.H.: Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc. Inf. Res. **19**(2), 79–85 (2013)
15. Metaxas, O., Dimitropoulos, H., Ioannidis, Y.: AITION: a scalable KDD platform for Big Data Healthcare. In: 2014 International Conference on IEEE-EMBS Biomedical and Health Informatics (BHI), pp. 601–604 (2014)

Big Data, Big Opportunities: Revenue Sources of Social Media Services Besides Advertising

Julian Bühler(✉), Aaron W. Baur, Markus Bick, and Jimin Shi

ESCP Europe Business School Berlin, Berlin, Germany
{jbuehler,abaur,mbick}@escpeurope.eu,
jimin.shi@edu.escpeurope.eu

Abstract. Facebook, Twitter, Instagram, and other players in the social media world have been on the rise during the last couple of years. In contrast to their popularity, their underlying business models are vague and often only linked to advertising. In this explorative study we identify new revenue sources for social media service providers besides advertising. Based on three use cases with Facebook, Tencent, and LinkedIn, we identify three possibly fruitful ways to extend existing social media business models. Subsequently, a survey with 301 respondents changes perspectives on the user's willingness to pay in order to identify usage-related differences evoked by cultural and external circumstances. Four derived hypotheses lead the way to avenues of further research especially in terms of Big Data analytics with new e-commerce trends like Facebook's Buy Button.

Keywords: Social media · Advertising · Business models · Big data · Case study

1 Introduction

On 30 January 2015, one of the world's leading social media services Facebook [1] updated its terms and data policy significantly by introducing new advertisement rules [2]. Users had to accept these changes in order to be able to still use the service. The modifications helped Facebook be more efficient in online behavioral advertising by collecting user information from various websites automatically, with the help of cookies. This behavior suggests that advertising is still one of the most important revenue pillars in the social media world. For instance, Facebook reported that income from advertising represented 88 % of its total revenue in Q2/2013 [3] and increased to 93 % in Q4/2014, compared to "payments and other fees revenue," comprising only 7 % [4]. However, tracking user reactions to this personalized advertising automatically leads to a massive overload of information which require modern analytic processes like certain Big Data algorithms. Still it remains unclear how and if Facebook interprets the tracked data, as not necessarily storage or analysis, but effective and efficient transformation of Big Data into reliable information causes problems [5].

In contrast to this development, Clemons [6] already forecasted six years ago that advertising would fail to play a role as the leading revenue source for Internet-based companies. The author suggests that consumers tend to ignore advertising including

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 183–199, 2015.
DOI: 10.1007/978-3-319-25013-7_15

online variations for three main reason: no trust in advertising, no willingness to see advertising, and no need for advertising at all to become informed [6]. The last argument is especially linked to the Internet as it has become easier for customers to collect all necessary information to decide on product purchases via search engines and product comparison websites. Therefore, it might be doubtful whether the focus on Big Data algorithms in advertising contexts is worth deeper research.

Since Clemons' analysis in 2009, the rise of social media has led to an unforeseen dominance of services including Facebook, Twitter, YouTube, and Instagram amongst others on the Internet. Thus, in this study we aim to identify existing alternative business models for revenue generation besides advertisement, as especially smaller or more specialized social media services struggle to capitalize on this single business model in the same way. The starting point for us to answer this question is a valuable categorization by Zambonini [7] that sheds light on other potential business models. However, social media was not considered explicitly despite the dominant position it has, nor did other promising studies recently (e.g., [8]). Until now, information systems (IS) or marketing research has not yet focused on operationalization of alternative revenue sources that trigger users' perceptions appropriately. In this empirical study we analyze different existing revenues, seeking to answer our main research question: *Which revenue sources besides advertising should social media service providers utilize?* Directly linked to this question are the underlying techniques in terms of suitable Big Data analytics routines [9].

The structure of this study is as follows. We first briefly discuss the term *social media* and its cognates as well as advertising in the context of this study (Sect. 2). Then we explain our research method, which contains three case studies of Facebook, Tencent, and LinkedIn and a survey with 301 participants (Sect. 3). Afterwards, we present three alternative revenue sources besides advertising based on the case study results and analyze them from the perspective of social media users (Sect. 4). We conclude this explorative study by deriving research hypotheses for accurate theory development in this field of study and with respect to Big Data analytics.

2 Conceptual Background

2.1 Definition of Business Model and Social Media Services

Before analyzing existing business models besides advertising in the social media environment, it is necessary to briefly clarify our understanding of these terms. We follow the definition introduced by Osterwalder and Pigneur [39, p. 14] who define a business model as 'the rationale of how an organization creates, delivers, and captures value". In addition and according to Kaplan and Haenlein [10, p. 61], 'social media services' can be understood as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content." The term 'social media services' is also linked to Web 2.0 by other authors and frequently used similarly to the expression 'social network' or 'social network site' (SNS) (e.g., [11–13]). We follow the first definition from Kaplan and Haenlein as it addresses the most relevant attributes of social media, but we consistently

use the term social media services instead of social network (sites) in this paper. In the context of Big Data infrastructure, social media services are an appropriate object of study as they can serve as a certain 'source type'.

2.2 Advertising as Financial Source for Social Media Services

Of 16 business models analyzed by Zambonini [7], advertising is the most common financial source for all web-based services and applications and can be classified as third-party supported revenue, which also includes social media services. Facebook as one of the largest social media services "had third quarter advertising revenue of $1.8 billion" [14, p. 1], being the largest share of overall revenues of $2.02 billion. Traditional online advertising is randomly displayed on the screen of a user, unrelated to the person's demographic characteristics, cultural background, or user preferences. While this type of advertising was used by many providers in the first years of their existence, only a few – including one of the largest microblogging services from China, Sina Weibo – are still using this advertising strategy.

A more common variation nowadays is user-related advertising. It is a kind of targeted advertising that exploits user data to personalize the ads shown with the goal of increasing the click through rate and the conversion rate. It entails elements of users' data (e.g., language and location) [15], browser history (by means of cookies saved in browsers), and social media activities. All of these social media related activities are tracked, recorded, and registered into a social graph, which provides advertisers with relevant data. However, a problematic "industry-driven obsession with the 'social graph'" [16] can occur if the underlying Big Data analysis procedures are too technical and lead to invalid interpretations of activities. Such activities may comprise the action of adding contacts, Facebook Likes on a page or post, as well as comments on other users' timelines and actions that involve online applications such as games, music, or news. Facebook's latest terms of January 2015 stretched the area of influence significantly, allowing the company also to retrieve remote information from other websites or digital services. Mobile advertising which is not explicitly addressed in this study adds the component of accessibility and local optimization to this concept. This enables location-based service functionality and, thus, timed place-sensitive advertisements such as notifications of sales and special events [17].

3 Research Method

3.1 Research Design: Case Study Analysis

We chose a case study approach because it is suitable for our explorative setting. According to Yin et al., "A case study is an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" [18, p. 13]. Following the meta-analysis on case study research by Dubé and Paré [19] for methodological procedure, we postulated a clear research question in the introduction. Based on our general social media service definition, we decided to apply a *multiple case design* to accurately address both the

diverse functionality of services and cultural diversity of users. All main elements, which are described in this and the next subchapter, are summarized in the complete visualization of our research model and approach (Table 1).

Table 1. Research model and approach

<table>
<tr><td colspan="3">Research Design</td></tr>
<tr><td>Clear Research Question [20–23]</td><td>Multiple Case Design [18, 20, 24]</td><td>Unit of Analysis [18]</td></tr>
<tr><td>“Which revenue sources besides advertising should social media service providers utilize?”</td><td>▪ Facebook
▪ Tencent
▪ LinkedIn</td><td>▪ Type A: entertainment-oriented social media services
▪ Type B: business-oriented social media service</td></tr>
<tr><td colspan="3">Data Collection</td></tr>
<tr><td colspan="3">Multiple Data Collection [18, 25]</td></tr>
<tr><td colspan="3">▪ Website analysis of three social media services (incl. terms of use, sitemaps, FAQs, and user accounts)
▪ Questionnaire addressing social media users and their willingness to pay for features, functions, or services</td></tr>
<tr><td colspan="3">Data Analysis</td></tr>
<tr><td colspan="2">Logical Chain of Evidence [18, 20]</td><td>Cross-Case Patterns [21, 23, 26]</td></tr>
<tr><td colspan="2">▪ Identification of relevant social media services (Step 1)
▪ Identification of existing revenue sources (premium features, services, etc.) besides advertising (Step 2)
▪ Analysis of users’ willingness to pay for these offers (Step 3)</td><td>▪ Similarities and differences of willingness to pay for offers for both social media service types A and B
▪ Empirical analyses as the basis</td></tr>
</table>

We used the Alexa website ranking [1] to identify the most globally used social media services and to distinguish between two types of social media services for our cases: entertainment-oriented social media services and business-oriented social media services. For us, entertainment-oriented social media services focus on general social networking tasks like picture sharing, pin boards, and gaming and emphasize predominantly leisure activities. On the other hand, we consider a business-oriented social media service as a service designed to share people’s skills and professional interests, and to promote oneself. Together, these two variations form our *unit of analysis* [19].

Following the Alexa ranking, we focused on the two leading entertainment services and the largest business service. First in the ranking is Facebook, which is not only the leading entertainment-oriented social media service according to our categorization, but also the most viewed website overall, second only to Google. In addition, we selected Tencent as a second entertainment service because it is present in a large, but partly restricted, market—China—and can reveal unknown phenomena. This cultural and legal distinction is in line with Yin’s definition of adequate case selection [18]. Tencent as a portal is also

present with two services in the Top 20, Qq.com and Weibo.com. Regarding the business-oriented social media services, we chose LinkedIn, which is the unchallenged worldwide leader and reaches the 13th rank in the Alexa ranking of all websites (not only social media services) [1].

3.2 Multiple Data Collection: Website Analysis and User Survey

According to Dubé and Paré's [19, p. 615] analysis of Sawyer [25] and Yin [18], "A major strength of case study data collection is the opportunity to use many different sources of evidence to provide a richer picture of the events". Hence, a *multiple data collection* approach was used as we first analyzed the three selected social media service websites with regard to offered revenue sources besides advertising. In addition, we developed a questionnaire to identify users' perceptions and adoption of these business models especially with respect to cultural and legal differences.

The general data collection phase was thus split into two parts (A & B) as well, starting with an exhaustive website analysis including terms of use and sitemaps (A). We additionally created user accounts on all three services to gain an overview of all offered functions and services that could potentially serve as alternative revenue sources for the service providers. After the selection of Facebook, Tencent, and LinkedIn (Step 1), this identification of revenue streams ties in afterwards as a second step within a *logical chain of evidence*. We then distributed a questionnaire to a non-specific target group of social media users in a second part of the data collection phase (B) to identify their willingness to pay for additional services and functions (Step 3).

The questionnaire itself is separated into three main sections and follows a funnel approach design [27], starting with questions on the demographic and socioeconomic situation of the participants. In the second section, we address the general usage of social media services and ask participants about their level of activity on associated social media websites. For all social media services they use with an individual account, conditional questions are designed to gather information on their willingness to pay for certain extra features, which we identified during the website analysis. The third section focuses on identifying social media services in professional or business surroundings in particular. We aimed to gain insights on user behavior here because we especially expected services such as LinkedIn to earn money with alternative business models besides advertising. We ask participants to give reasons why they use these services and which types of premium functions, upgrades, or other functionalities are subject to costs. Both for entertainment- and business-oriented services, participants are also requested to state prices they expected to pay for these extra features. This allows us to identify *cross-case patterns* between the three case studies. The final part of our questionnaire addresses how users intend to use these services in the future.

As this is an explorative approach, we have the goal of making inferences on a broad and random sample of the population and do not define a concrete target group for this study. Thus, possible exclusions of participants due to potential group bias issues can be avoided *a priori* [27, 28]. The only characteristic participants have to fulfill is that they are Internet users and/or have Internet access regularly. This single criterion is helpful to address participants who are already online and probably active on or familiar

with social media services. In line with the population we strive for, i.e., Internet users in general, we designed our questionnaire to be digital, not paper-based. Even though a few users might not be able to access the survey without hindrance due to technical problems or lack of support [27, 28], it is an appropriate distribution means in our context. Therefore, the link to the corresponding survey was spread through various channels. This includes several mailing lists, social media services, and instant messengers as well as offline notices with short links and QR codes to our survey at many physical locations like universities and stations, being accessed by heterogeneous visitors.

4 Analysis and Results

4.1 Case Studies: Identification of Alternative Financing Strategies

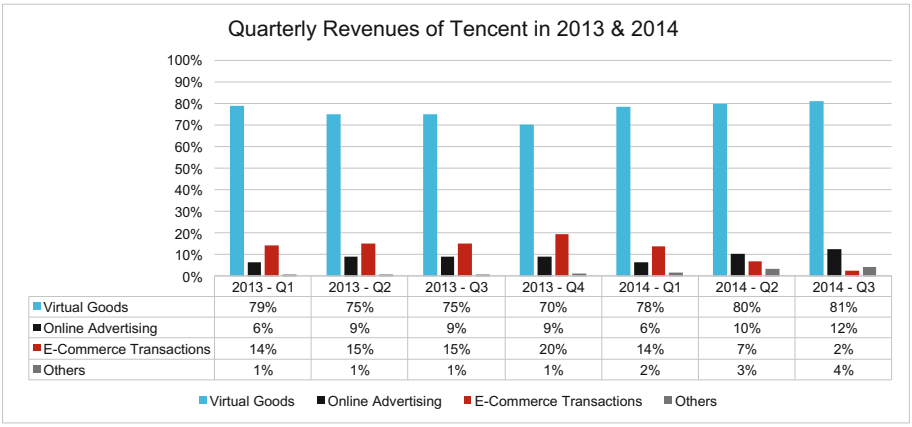
Case Study 1: Virtual Goods on Tencent. In our first case study, we focused on one of the largest social media service providers worldwide, the Chinese company *Tencent*. It offers several services via its platform qq.com (Alexa rank 10) like China largest social networking site similar to Facebook called *Q Zone*, a microblogging service similar to Twitter called *Weibo*, and other services like instant messaging (*WeChat*) or games (*QQ Game*). We decided to analyze Tencent besides its high ranking as it represents a large target group of Asian social media users as well as a broad variety of services and functions.

We analyzed the website including subpages and logged in as regular users to receive insights from their perspective. Besides advertising, we initially learned that Tencent has created thousands of different virtual goods for their numerous services. These virtual goods are offered to users for money and can be seen as "intangible objects purchased in order to be used in online communities such as SNS sites or online games. They also comprise virtual money (or virtual currency) which is used to purchase these intangible or physical goods" [29]. Ho and Wu [30, p. 208] investigated user's intention to purchase virtual goods especially in online games and according to the authors they became "a major source of income" for service providers. As especially the "social dimension […] is likely to influence the purchasing intentions of individuals" [31, p. 790], virtual goods can play an important role not only for online games but social media service providers in general.

From what we found, the virtual goods offered by Tencent can be divided into two categories by their functions. The site offers goods with customization purposes, such as wallpapers and decorations for personal space. Virtual goods of this category are mostly offered on the social networking service Q Zone, but can also be found on WeChat in terms of emoticon stickers for chatting or outfits and cosmetics for customizing one's avatar. The second category consists of virtual goods offered for gaming purposes, such as tools and credits which users can utilize in certain games. According to Guo and Barnes [32], this is both an emerging trend for South Korea and China as Asia's biggest gaming markets as well as the USA (currently the largest e-gaming market). Unlike other social media services, e.g., Facebook, Tencent does not open itself as a platform for third party developers to offer their own games but soleily develops its own game.

After analyzing the offered types of virtual goods, it is important to clarify how Tencent is actually earning money with them. The company uses its own virtual money called Q coin which users can buy to purchase the virtual goods. The value of Q coin is connected to the Chinese currency Renminbi (CNY) and 10 Q Coins are worth 10 CNY (or ≈ 1.6 USD). This virtual money can be bought via different channels, such as QQ's official online payment site, mobile credits, post offices, convenient shops, or even newspaper stands around China. Purchased this way, Q coin serves like a prepaid card that users need to purchase in advance to deposit the equivalent money into their account. A second option is offered by Tenpay, which is a third party payment platform launched by Tencent connected with users' debit/credit card account or other online payment solutions such as PayPal. Once bought, users can share the same account information to purchase virtual goods with Q Coins among all the social media services of Tencent.

From this case study analysis we summarize that Tencent follows a strategy which allows its users to have free registration in order to maximize users' acquisition. Afterwards Tencent starts to push its charged virtual goods to the users which upgrade their accounts or improve their social media experience. This paid upgrade of user accounts to a premium one, which entails the possibility to exploit several different features precluded to the free accounts, is the main characteristic of the popular business model denominated "freemium" [33]. The term "freemium" describes a business model which combines "'free' and 'premium' consumption in association with a product or service" [34, p. 1]. Figure 1 shows that unlike most other social media service providers, the biggest proportion of Tencent's revenues is not advertising but virtual goods. Sold with the help of the freemium concept, they had a share between 70 % (minimum, Q4/2013) and of 81 % (maximum, Q3/2014) over the total revenues in the last two years [35]. These results suggest that Tencent was successful in establishing a financing method for its social media service besides advertising.



	2013 - Q1	2013 - Q2	2013 - Q3	2013 - Q4	2014 - Q1	2014 - Q2	2014 - Q3
Virtual Goods	79%	75%	75%	70%	78%	80%	81%
Online Advertising	6%	9%	9%	9%	6%	10%	12%
E-Commerce Transactions	14%	15%	15%	20%	14%	7%	2%
Others	1%	1%	1%	1%	2%	3%	4%

Fig. 1. Comparison of revenue sources for Tencent in 2013 & 2014 (Source: [35])

Case Study 2: e-commerce and Social Commerce Services on Facebook and Tencent. Social media services can also provide users with e-commerce functions within their environment, for example online trading and payment services. Functionalities including transactions between businesses and customers (B2C, as well as B2B and C2C) are often offered on the website of the social media service or as a third party payment platform on other websites. We analyze these activities which are sometimes referred to as social commerce in our second case study.

In the context of e-commerce functionality, Facebook seems to follow a strategy of trying out several approaches very fast according to our analyses. During the last couple of years, Facebook first had started with a service called Credits. Facebook Credits was a payment service for its platform of games and apps, developed and offered by third party developers. Users could purchase Facebook Credits and then convert them automatically into the in-game or in-app items they need. Facebook served as an online payment service and all purchases of virtual goods for applications and games on Facebook's platform were processed through Facebook Credits, the only payment option [36]. This service started in 2011 and developers earned 70 % of the value while Facebook retained 30 %. It was later replaced in 2013 by a service called Local Currency Payments, a service that converts the value of a virtual good into local currencies of users. The most common payment channels are credit card/debit cards, PayPal and mobile technologies (e.g., Google Wallet), but alternative local payment methods exist, varying from country to country [37]. For example, Moneybookers is available in the U.S. and Giropay in Germany. The service fee charged from game and app developers remained unchanged. The latest changes include Facebook Gifts, an online trading platform offered by Facebook in September 2012, where users could buy physical 3rd party company gifts or vouchers for their friends. This service was closed in 2014 and at the moment, users can only buy vouchers in the form of game cards in some countries. After various tests, Facebook will most likely introduce a new e-commerce feature called Buy Button in 2015, which allows third party companies to publish posts directly via the Facebook profile. The buying process for users is supposed to be easier as purchases can be made without leaving Facebook's website.

Besides the virtual goods business described in the first case study, our website analyses reveal that Tencent also play an important role in China's e-commerce sector. There are two individual services, PaiPai.com and Tenpay, which both are connected with Tencent's social media services like Q Zone. According to Tencent, PaiPai.com is a B2C & C2C online trading platform launched in 2006 where users can find products from different categories, for example, clothing, electronic or education products [38]. They can log-in with their account from Tencent and every purchase can be shared among the various services. The business model of PaiPai.com is not based on account fees, which do not exist for sellers, but on added values. Sellers can buy trust certificates and receive in return e.g. extra space for larger product pictures in high resolution. The second e-commerce service offered by Tencent is Tenpay, which can not only be used to buy Q Coins, but also for regular bank transactions like online payments, money transfers, or others. Unlike Facebook's payment methods, Tenpay is a third party payment method available not only for Tencent's social media services, but also for the platforms of other companies.

Results of this case study with both companies reveal that major social media services are currently active in the e-commerce sector and even though Facebook changed its strategy over the years, e-commerce seems to be a promising alternative financing strategy to advertising.

Case Study 3: Account Upgrade Services on LinkedIn. In our third case study, we analyze LinkedIn[1] as the world's largest representative of other business-orientated social media services like Xing or Viadeo. LinkedIn offers a previously described freemium business model including four types of certain premium and account upgrade services. The first one is named *job seeker* and is designed for individual users looking for a new job opportunity. The second option is labeled as *business plus*, where experienced LinkedIn users can use extra functionality, such as additional search features. *Sales plus* is supposed to be used by salespeople looking for new leads or business partners and *Recruiter Lite* is dedicated to company representatives seeking for new employees via the LinkedIn platform. In addition to these four premium subscriptions, LinkedIn offers two additional premium services called *Talent Solutions* and *Marketing Solutions*. Talent Solutions is designed to support larger companies during their recruiting process while Marketing Solutions allows customers to present their content in a prominent position both on LinkedIn's website and the mobile app. Once users subscribe to one of the upgrade options, they get access to features included in the premium package.

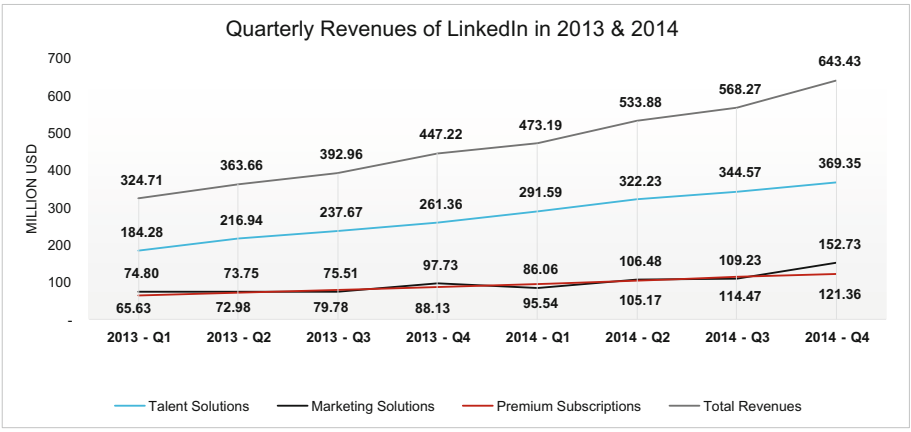


Fig. 2. Comparison of different revenue sources for LinkedIn in 2013 & 2014 (Source: LinkedIn revenue reports, http://investors.linkedin.com/results.cfm)

We consolidated LinkedIn's information of the quarterly revenue reports[2] and Fig. 2 illustrates the revenue sources for 2013 & 2014. They indicate a successful adop-

[1] Service description of LinkedIn: https://www.linkedin.com/about-us.

[2] LinkedIn: Quarterly Earnings 2013/2014: http://investors.linkedin.com/results.cfm?Quarter=&Year=2013 and http://investors.linkedin.com/results.cfm?Quarter=&Year=2014.

tion of the freemium concept with steadily growing revenue over the years. Anticipating similar results for 2015 following the recent trend, we consider this business model being a potential alternative financing strategy to advertising for social media service providers as well. The business model *advertising* was previously categorized as third-party supported. In contrast, the analyzed business model account *upgrade services* comes within limits of subscription, pay for additional premium content or advanced features. Both *virtual goods* (here: virtual products) and *e-commerce* belong to payments according to [7]. All three alternative models identified in the case studies build the underlying structure for our second empirical part of this paper, a survey which includes participants in the core target markets of social media services, China, the USA, and Europe.

4.2 Survey Results

Demographic Structure. After a period of six weeks we received 333 responses to our online survey in total. Out of these participants, 31 did not fully complete the survey and were excluded from our sample and further analyses. One additional participant was removed from our sample due to insufficient data reliability, which results in a final sample size of n = 301 respondents. The demographic distribution of our final sample is illustrated in Table 2, revealing non-representativeness but wide range regarding the age and varying nationalities of the participants, as strived for:

Table 2. Demographic distribution of survey respondents

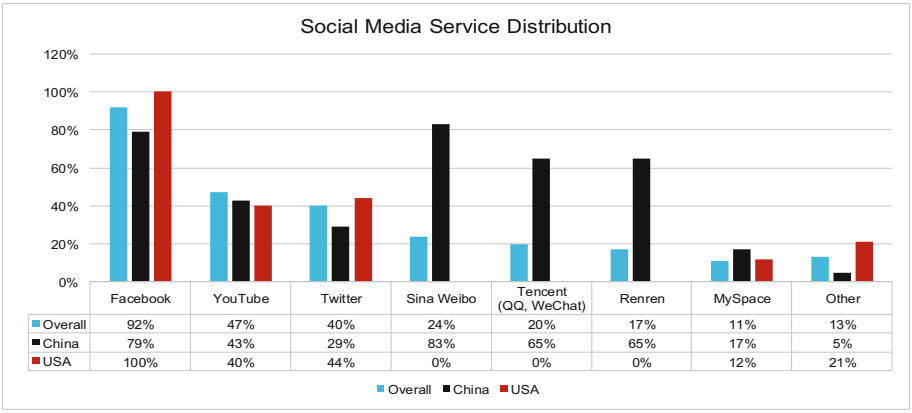
Demographic category	Distribution
Age	33.7 years (sd: 11.1)
Gender	61.1 % female, 38.9 % male
Nationality	20.27 % from North America (incl. 14.95 % from the US, 4.64 % from Canada)
	31.23 % from Asia (incl. 26.25 % from China)
	45.18 % from Europe (incl. 13.62 % from Germany, 6.98 % from Italy, 6.31 % from France)
	03.32 % from other continents

Another important characteristic of the sample of respondents is their status of employment. The majority of participants stated they are employed for wages (44.19 %), followed by 21.93 % who are students and 12.62 % who are interns. The remaining participants are self-employed (8.31 %), currently looking for a job (5.32 %), retired from work (3.65 %), or in other employment (3.99 %). This sample structure is relevant especially with respect to the sections of the questionnaire dealing with willingness to pay. More than half of the participants—at least employees for wages and self-employed people—can be considered to earn money

with their job and, therefore, have the purchasing power to pay for virtual goods or premium functionality.

Entertainment-Oriented Social Media Usage. The popularity of social media usage among our survey participants is clearly visible according to the percentage of people with a subscription and account for at least one of the entertainment-oriented services. We provided predefined answers for seven important social media services, but respondents could name additional services in a text box. Results show that 92.69 % (n = 279) of the respondents have subscribed to a social media service and an additional 2.33 % is planning to do so. Only less than five percent (4.98 %) of the sample is not interested in using social media services at all, right now, or in the future. Overall, results indicate a high level of activity as 265 of the 279 social media users in our questionnaire use the services at least once a week, 98.61 % of them even once or several times a day. These results were measured on a 7-point Likert scale ranging from –3 ("I never use social media service") to +3 ("I use social media service a couple of times a day").

Going deeper into the analysis, the distribution of subscriptions to different social media services throughout our sample shows the clear supremacy of Facebook, followed by YouTube and Twitter. These results are in line with common rankings based on page impressions, e.g., the Alexa ranking, and confirm the dominant position occupied by this services globally. A more detailed look results from splitting the sample group into geographical clusters by continent membership. While Facebook reaches a saturation of more than 73 % in all continents including Asia, even though access is technically restricted in China, other services vary significantly. As anticipated, all three Chinese-based social media services—Renren, Sina Weibo, and Tencent—are used by nearly 50 % or more users in Asia but are virtually not recognized by those coming from North America or Europe. Figure 3 gives an overview of the saturation of main social media services by continent.



	Facebook	YouTube	Twitter	Sina Weibo	Tencent (QQ, WeChat)	Renren	MySpace	Other
Overall	92%	47%	40%	24%	20%	17%	11%	13%
China	79%	43%	29%	83%	65%	65%	17%	5%
USA	100%	40%	44%	0%	0%	0%	12%	21%

Fig. 3. Social media service distribution in China, the US, and worldwide (Source: own survey results)

Although entertainment-oriented social media services are being used quite heavily, our results indicate that purchases are relatively rare in this field. Only 31 respondents (10.80 %) bought additional functions or products. Age differences are not noticeable, but regarding the continent-based distribution, results clearly show that Chinese survey participants are the majority who purchased virtual products including virtual money on social media services. In our survey setting, participants were provided with certain offers like virtual goods we identified within the case studies, but they could also name other services or virtual goods. Respondents from Asia paid predominantly for premium accounts (e.g., Sina Weibo VIP accounts; Tencent QQ diamond account), virtual products (e.g., Tencent QQ avatar outfits, pets, and new themes; Tencent WeChat emoticon stickers), and virtual money (e.g., Tencent Q Coins). On the other hand, only survey participants from the US purchased Facebook gift cards, while Europeans had no particular key feature. On average, these 31 respondents spent $21.84 for the additional functions and products (Fig. 4). These results clearly indicate a cultural difference in terms of usage interests between Asian (in particular Chinese) and European or North American social media users.

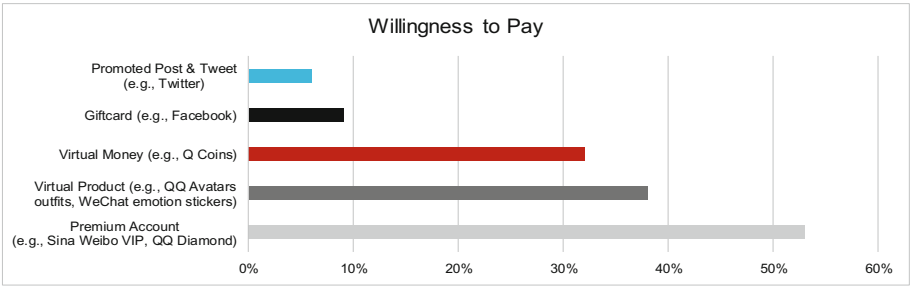


Fig. 4. Willingness to pay for selected functions and products (Source: own survey results)

The last part of this second section of our questionnaire addressed the willingness to pay money for additional functions and products. Overall, social media services in this field have low potential here according to our results because only 3.97 % of existing and prospective users stated their interest in spending money for these features on entertainment-oriented social media services. Of all offers, virtual products in the Asian market have the highest potential compared to others.

Business-Oriented Social Media Usage. The final part of our questionnaire addressed business-oriented social media services. We provided the respondents with ten predefined answer options according to the ten most used services in Alexa ranking (Absolventa, Biznik, Cofoundr, Ecademy, E.Factor, LinkedIn, Ryze, Tianji, Xing, and Ziggs) and included an additional option to name "other service." The overall percentage of respondents using these services is significantly lower compared to entertainment-oriented services, but still 64.78 % (n = 195) subscribe to one and 14.29 % (n = 43) plan to subscribe in the future. LinkedIn is the unchallenged leading service in this field with 168 users (55.81 %) within our total sample and 24 (7.97 %) of them from China. This is interesting as LinkedIn released its beta version of a simplified Chinese site only in

February 2014 and a large new target group of Chinese-speaking members potentially joined the service since then. Other services than LinkedIn, in contrast to entertainment-oriented social media services, only play a role in individual regions and countries according to our results. For example, XING, a business network with similar functionality as LinkedIn, is used by 61 participants (20.27 %) even though it is only available in the German-speaking market and Spain. Viadeo was named by only three participants.

The next block of survey questions aimed at identifying reasons why users tend to use business-oriented social media services. Among the many possible reasons for subscription, three seem to be the most popular among our survey participants. Throughout all continents and countries, "networking with professionals" (72.82 %) is extremely important to our survey participants (Fig. 5). Business-oriented services in the social media world have given a completely new playground to the professional networking experience, so that many of our survey participants claim to use their accounts to find new professional contacts and share views, ideas, and know-how, alongside a new name in their contact lists. A second reason for 56.92 % in our sample is related to the recruiting purpose of such services, which is to "find a job." A detailed look at this option reveals that together with students (18.97 %) and interns (16.41 %), the survey participants who use a social media service like LinkedIn to find a job are mostly employed for wages (50.29 %). These results suggest that LinkedIn and others are nowadays an important tool used by employed people to steer their careers and to find new job opportunities. Additionally, visibility within professional communities ("Become visible active in professional communities," 51.28 %) as the third important option supports this line of argumentation.

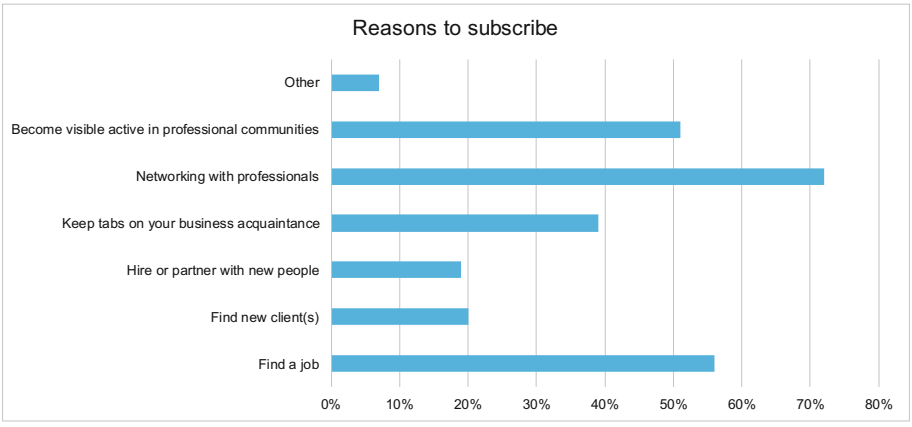


Fig. 5. Reasons to subscribe to business-oriented social media services (Source: own survey results)

These findings directly lead to the last part of this survey, which addresses the willingness to pay for premium functions in business-oriented social media services. Overall, 12.82 % (n = 25) of the 195 users of these services stated that they already paid

for premium features, and many of them confirmed in an open question design that gaining access to more features and functions in order to get in contact with new business partners was the main reason. Especially LinkedIn's direct mail function called InMail, which comes along with all payment models—analyzed in the third case study—seems to be a key function users are willing to pay for. Additionally, extended search functions including new algorithms (offered by, e.g., LinkedIn and XING) are important for both job seeking users and recruiters from companies. According to our results, the majority of the current users of such premium features (88.00 %) is satisfied with them and will keep the subscription.

Our survey participants could state in a text box the amount of money they are potentially willing to pay per month in total for business-oriented social media services. The answer range is quite broad, from a minimum of $1.81 to a maximum of $113.00. A distinct pattern with regard to willingness to pay does not exist according to these findings, since the results show a pretty high standard deviation ($21.04). But the average value of $17.00 for all respondents, including potentially new members of these services, is comparable to the more frequently used entertainment-oriented social media services.

5 Conclusion

5.1 Findings and Hypotheses

With explorative case studies, we first identified three alternative business models, i.e., *virtual goods*, *e-commerce*, and *account upgrade services*. This extends the more general research by Zambonini [7] and Vukanovic [8]. We then analyzed the current user behavior on prominent entertainment- and business-oriented social media services. A clear focus was set in this second empirical part on questions directly related to willingness to pay for these extra functions or products. Results indicate that alternatives exist to challenge the cash cow advertising. From our survey and our second case study we confirmed that virtual goods are particularly used for customization and gaming purposes, e.g., on Facebook or Tencent's Q Zone. On the other hand, account upgrades are bought by users of both types of social media services analyzed in this study, but predominantly on business-oriented ones like LinkedIn. Therefore we derive the following hypotheses, which can be tested in confirmative follow-up studies:

H1: *Users of entertainment-oriented social media services tend to buy virtual goods while those of business-oriented social media services spend money on account-upgrade offers.*

We could also identify differences in the use behavior influenced by what we think can be seen as three moderators: geographical location (H2), age (H2, H4), and availability of a service in a certain language (H3). These moderators change the relationship postulated in H1:

H2: *Asian users of entertainment-oriented social media services, especially younger ones, tend to buy more virtual goods and less account upgrades compared to Europeans and North Americans.*

H3: *Availability in a native language moderates the effect of users striving for business-oriented social media services for all users. The more languages offered by a business-oriented service provider, the more users will join the service.*
H4: *Age moderates the effect of striving for social media presence for all users. The older a user grows, the more he or she tends to be present in business-oriented social media services.*

Business-oriented social media services are less present in Asia compared to Europe and North America, but this market penetration seems to have changed rapidly since the market leader LinkedIn made its services available in both the Japanese and Chinese languages. We think that this decision will boost memberships accompanied by a high potential of selling account upgrade services. With regard to our initially postulated research question, providers of entertainment-oriented social media services should concentrate on virtual goods in the gaming sector, especially with particular focus on Asian users. According to our results, they are open minded to alternative services and can set a trend, which could then be followed by European and North American users. In contrast, business-oriented service providers should utilize premium functions and account features as part of the well-established freemium concept.

5.2 Limitations and Further Research

In general, we could prove that users are willing to pay for additional features within social media services. However, this environment is absolutely fast moving with new ideas, functions, and implementations aspiring daily, like Facebook being on the brink of introducing its Buy Button after trying out several precursors. We had to set limits on this initial explorative approach, and we think that in addition to leading players like Facebook, Tencent, and LinkedIn, more and new emerging social media services like Instagram, which struggles to make money at the moment, should be examined as well.

From the empirical data perspective, we gained a suitable sample size of 301 respondents for this explorative setting. The hypotheses deduced are based on data from individuals from all around the leading social media markets in North America, Asia, and Europe. This already covers a decent percentage of general cultural clusters. In upcoming studies, however, we suggest researchers narrow the scope of their empirical design to one of the core markets in order to shed light on specific cultural characteristics. We think further research with individual analyses, especially on users' willingness to spend money for extra services in these markets, could provide new insights.

The initial part of this research paper described Facebook's radical move of introducing new advertisement rules at the beginning of this year. Rather than being a gift for its users, this leading company in the social media environment consequently took the next step on the way to optimized user-related advertising. Right now advertising is by far the most profitable financing strategy of social media services. However, it is not the only effective one, as some players like Tencent demonstrate. Advertising is not the whole story when it comes to revenues, and new ways like virtual goods or freemium concepts are on the horizon. In the end, convenient e-commerce functionalities offered directly within a social media service—like Facebook's Buy Button—could be

a condign revenue competitor to advertising. But regardless of the revenue source used, tracking the activities will be of utmost importance. Social media service providers should have this in mind, especially as "managing context in light of Big Data will be an ongoing challenge." [16, p. 671].

References

1. Alexa Internet Inc.: The top 500 sites on the web. http://www.alexa.com/topsites
2. Facebook: updating our terms and policies: helping you understand how facebook works and how to control your information. https://www.facebook.com/about/terms-updates
3. Melanson, D.: Facebook reports $1.81 billion in revenue for Q2 2013, 1.15 billion monthly active users. http://www.engadget.com/2013/07/24/facebook-q2-2013-earnings/
4. Statista: Facebook's global revenue as of 1st quarter 2015, by segment (in million U.S. dollars). http://www.statista.com/statistics/277963/facebooks-quarterly-global-revenue-by-segment/
5. Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., Ghosh, P.: Big data: prospects and challenges. J. Decis. Makers **40**, 74–96 (2015)
6. Clemons, E.K.: Business models for monetizing internet applications and web sites: experience, theory, and predictions. J. Manag. Inf. Syst. **26**, 15–41 (2009)
7. Zambonini, D.: Monetizing your web app: business model options. http://www.boxuk.com/blog/monetizing-your-web-app-business-models/
8. Vukanovic, Z.: New media business models in social and web media. J. Media Bus. Stud. **8**, 51–67 (2011)
9. Goes, P.B.: Big data and IS research. MIS Q. **38**, iii–viii (2014)
10. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of social media. Bus. Horiz. **53**, 59–68 (2010)
11. Chiang, I., Huang, C., Huang, C.: Characterizing web users' degree of web 2.0-ness. J. Am. Soc. Inform. Sci. Technol. **60**, 1349–1357 (2009)
12. boyd, d., Ellison, N.B.: Social network sites: definition, history, and scholarship. J. Comput.-Mediated Commun. **13**, 210–230 (2007)
13. Maier, C., Laumer, S., Weitzel, T.: Although I am stressed, I still use IT! Theorizing the decisive impact of strain and addiction of social network site users in post-acceptance theory. In: Proceedings of 34th International Conference on Information Systems, Milan, Italy (2013)
14. Wilhelm, A.: Facebook's desktop ad revenues fell $26 m, in Q3 as its mobile ad revenue surged $226 m. http://techcrunch.com/2013/10/30/facebooks-desktop-ad-revenues-fell-26m-in-q3-as-its-mobile-ad-revenue-surged-226m/
15. Gatto, K.: How facebook sells your personal information. http://news.discovery.com/tech/gear-and-gadgets/how-facebook-sells-your-personal-information-1301241.htm
16. boyd, d., Crawford, K.: Critical questions for big data. Inf. Commun. Soc. **15**, 662–679 (2012)
17. Johnston, K.: What are the advantages of mobile advertising? http://smallbusiness.chron.com/advantages-mobile-advertising-55314.html
18. Yin, R.K.: Case Study Research. Design and Methods. Sage Publications, Thousand Oaks (1994)
19. Dubé, L., Paré, G.: Rigor in information systems positivist case research: current practices, trends, and recommendations. MIS Q. **27**, 597–635 (2003)
20. Benbasat, I., Goldstein, D.K., Mead, M.: The case research strategy in studies of information systems. MIS Q. **11**, 369–386 (1987)

21. Eisenhardt, K.M.: Building theories from case study research. Acad. Manag. Rev. **14**, 532–550 (1989)
22. Mays, N., Pope, C.: Qualitative research: rigour and qualitative research. BMJ **311**, 109–112 (1995)
23. Miles, M.B., Huberman, A.M.: Qualitative Data Analysis. An Expanded Sourcebook. Sage Publications, Thousand Oaks (1994)
24. Lee, A.S.: A scientific methodology for MIS case studies. MIS Q. **13**, 33–50 (1989)
25. Sawyer, S.: Analysis by long walk: some approaches to the synthesis of multiple sources of evidence. In: Trauth, E.M. (ed.) Qualitative Research in IS. Issues and Trends. Idea Group Publishing, Hershey (2001)
26. Nisbett, R., Ross, L.: Human Inference. Strategies and Shortcomings of Social Judgment. N.J. Prentice-Hall XVI, Englewood Cliffs (1980)
27. Hair, J.F.: Multivariate Data Analysis. Prentice Hall, Upper Saddle River (2010)
28. Stopher, P.R.: Collecting, Managing, and Assessing Data using Sample Surveys. Cambridge University Press, Cambridge (2012)
29. Javelin strategy and research: virtual currency and social network payments – the new gold rush: how emerging virtual transactions will alter the payments landscape forever (2011). https://www.javelinstrategy.com/brochure/212
30. Ho, C.-H., Wu, T.-Y.: Factors affecting intent to purchase virtual goods in online games. Int. J. Electron. Bus. Manag. **10**, 204–212 (2012)
31. Animesh, A., Alain, P., Sung-Byung, Y., Wonseok, O.: An odyssey into virtual worlds: exploring the impacts of technological and spatial environments on intention to purchase virtual products. MIS Q. **35**, 789–810 (2011)
32. Guo, Y., Barnes, S.: Virtual item purchase behavior in virtual worlds: an exploratory investigation. Electron. Commer. Res. **9**, 77–96 (2009)
33. Anderson, C.: Free: the past and future of a radical price. Hyperion, New York (2009)
34. Niculescu, M.F., Wu, D.J.: When should software firms commercialize new products via freemium business models? (2011) http://rady.ucsd.edu/faculty/seminars/papers/Niculescu.pdf
35. Tencent: tencent announces 2014 third quarter and interim results. http://www.tencent.com/en-us/at/pr/2014.shtml
36. Eldon, E.: Facebook sets July, 1, 2011 deadline to make credits sole canvas game payment option. http://www.adweek.com/socialtimes/facebook-sets-July-1-2011-deadline-to-make-credits-sole-canvas-game-payment-option/255581?red=if
37. Facebook: facebook payments. https://developers.facebook.com/docs/payments/overview
38. Tencent: tencent - products and services – ecommerce. http://www.tencent.com/en-us/ps/ecommerce.shtml
39. Osterwalder, A., Pigneur, Y.: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Wiley, New Jersey (2013)

Big Data Analytics as a Service for Business Intelligence

Zhaohao Sun[1(✉)], Huasheng Zou[2], and Kenneth Strang[3]

[1] Department of Business Studies, PNG University of Technology, Lae 411, Papua New Guinea
zsun@dbs.unitech.ac.pg, zhaohao.sun@gmail.com
[2] College of e-Commerce, Ningbo Dahongying University, Ningbo 315175, Zhejiang, China
zoufan99@163.com
[3] School of Business and Economics, State University of New York, Plattsburgh at Queensbury, NY 12804, USA
kenneth.strang@plattsburgh.edu

Abstract. This paper proposes an ontology of big data analytics and examines how to enhance business intelligence through big data analytics as a service by presenting a big data analytics services-oriented architecture (BASOA), and applying BASOA to business intelligence, where our surveyed data analysis showed that the proposed BASOA is viable for developing business intelligence and enterprise information systems. This paper also discusses the interrelationship between business intelligence and big data analytics. The proposed approach in this paper might facilitate the research and development of business analytics, big data analytics, and business intelligence as well as intelligent agents.

Keywords: Big data analytics · e-commerce · Business intelligence · Intelligent agents

1 Introduction

Big data and big data analytics has become one of the important research frontiers [1]. Big data and its emerging technologies including big data analytics have been not only making big changes in the way the e-commerce and e-services operate but also making traditional data analytics and business analytics bring new big opportunities for academia and enterprises [2]. Big data analytics is an emerging big data technology, and has become a mainstream market adopted broadly across industries, organizations, and geographic regions and among individuals to facilitate data-driven decision making for business and individual's hedonism [3, 4].

Business intelligence (BI) has received widespread attention in academia, e-commerce, and business over the past two decades [5]. BI has become not only an important technology for improving business performance of enterprises but also an impetus for developing e-commerce and e-services [15]. However, BI is facing new challenges and opportunities because of dramatic development of big data and big data technologies [6, 7]; that is, how to use big data analytics to enhance BI becomes a big issue for business, e-commerce, e-services, and information systems.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 200–211, 2015.
DOI: 10.1007/978-3-319-25013-7_16

Big data analytics and BI are the top priorities of chief information officers (CIOs) and comprise a $12.2 billion market [8]. According to a study of Gartner, worldwide BI and analytics software, consisting of BI platforms, analytic applications and advanced analytics, totalled $14.4 billion in 2013, an 8 % increase from 2012 revenue [9]. This fact attracts unprecedented interest and adoption of big data analytics. According to the annual survey results of 850 CEO and other C-level executives of global organisations, McKinsey [2] concludes that 45 % of executives put "big data and advanced analytics" as the first three strategic priorities in both strategy and spending in three years' time and more than one thirds of executives will now spend or in three years' time in this area. IDC (International Data Corporation) predicts that the business analytics software market will grow at a 9.7 % compound annual growth rate over the next five years from 2012 to 2017 [4].

The above brief discussion and literature review implies that there is a close relationship between big data analytics and BI. However, the following two important issues have not been drawn significant attention in the scholarly peer-reviewed literature:

- What is the relationship between big data analytics and BI?
- How can big data analytics enhance BI?

This paper will address these two issues through extending our early research on analytics service oriented architecture [2]. To address the first issue, we propose an ontology of big data analytics in Sect. 2 through overviewing our early work on data analytics and big data analytics. To address the second issue, we examine big data analytics as a technology for supporting BI through examining the relationship between big data analytics and BI in Sect. 3. We then present a big data analytics service oriented architecture (BASOA), in which we also explore how to apply big data analytics as a service to enhance BI, where we show that the proposed BASOA is viable for developing BI based on our surveyed data analysis.

The remainder of this paper is organized as follows. Section 2 looks at the fundamentals of big data analytics by proposing an ontology of big data analytics and discussing the relationships of big data analytics and data analytics. Section 3 discusses BI and its relationships with big data analytics. Section 4 presents BASOA. Section 5 applies proposed BASOA to BI. The final sections discuss the related work and end this paper with some concluding remarks and future work.

2 Fundamentals of Big Data Analytics

This section proposes an ontology of big data analytics and looks at the interrelationship between big data analytics and data analytics. To begin with, this section first examines the fundamentals of big data analytics.

Big data analytics is an integrated form of data analytics and web analytics for big data [2]. According to [7, 10], big data analytics can be defined as the process of collecting, organizing and analyzing big data to discover patterns, knowledge, and intelligence as well as other information within the big data. Similarly, big data analytics can be defined as techniques used to analyze and acquire knowledge and intelligence from

big data [7]. Big data analytics is an emerging science and technology involving the multidisciplinary state-of-art information and communication technology (ICT), mathematics, operations research (OR), machine learning (ML), and decision sciences for big data [1, 2]. The main components of big data analytics include big data descriptive analytics, big data predictive analytics and big data prescriptive analytics [11].

- Big data descriptive analytics is descriptive analytics for big data [12], and is used to discover and explain the characteristics of entities and relationships among entities within the existing big data [13, p. 611]. It addresses the problems such as what happened, and when, as well as what is happening. For example, web analytics for pay-per-click or email marketing data belongs to big data descriptive analytics [14].
- Big data predicative analytics is predicative analytics for big data, which focuses on forecasting trends by addressing the problems such as what will happen, what is likely to happen and why it will happen [12, 15]. Big data predicative analytics is used to create models to predict future outcomes or events based on the existing big data [13, p. 611]. For example, big data predicative analytics can be used to predict where might be the next attack target of terrorists.
- Big data prescriptive analytics is prescriptive analytics for big data, which addresses the problems such as what we should do, why we should do it and what should happen with the best outcome under uncertainty [11, p. 5]. For example, big data prescriptive analytics can be used to provide an optimal marketing strategy for an e-commerce company.

An ontology is a formal naming and definition of a number of concepts and their interrelationships that really or fundamentally exist for a particular domain of discourse [16]. Then, an ontology of big data analytics is an investigation into a number of concepts and their interrelationships that fundamentally exist for big data analytics. Based on the above discussion, we propose an ontology of big data analytics, as illustrated in Fig. 1. In this ontology, big data analytics is at the top while big data and data analytics are at the bottom. Big data descriptive analytics, big data predictive analytics, and big data prescriptive analytics are at the middle level as the core parts of any big data analytics.

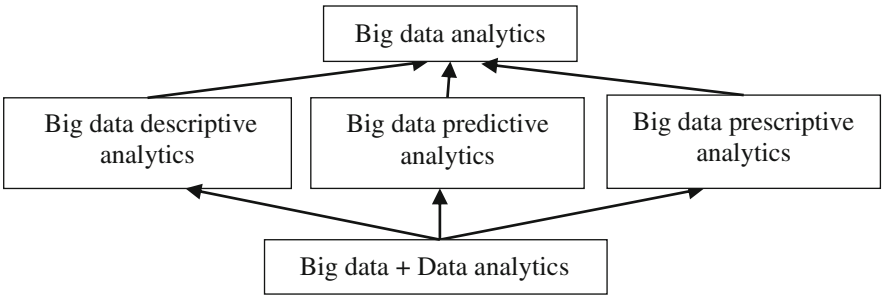


Fig. 1. An ontology of big data analytics

It should be noted that the above-proposed ontology of big data analytics is still simple. We will extend it by adding another level between the second level and the third

level in Fig. 1. This added level will elaborate big data descriptive, predictive and prescriptive analytics taking into account the corresponding real-world examples, methods and techniques.

In Fig. 1, data analytics refers to as a method or technique that uses data, information, and knowledge to learn, describe and predict something [15, p. 341]. In brief, data analytics can be then considered as data-driven discoveries of knowledge, intelligence and communications [12]. More generally, data analytics is a science and technology about examining, summarizing, and drawing conclusions from data to learn, describe and predict something [2].

The fundamentals of big data analytics consists of mathematics, statistics, engineering, human interface, computer science and information technology [1, 2]. The techniques for big data analytics encompass a wide range of mathematical, statistical, and modeling techniques [13, p. 590]. Big data analytics always involves historical or current data and visualization [17]. This requires big data analytics to use data mining (DM) to discover knowledge from a data warehouse (DW) or a big dataset in order to support decision making, in particular in the text of big business and management [15, p. 344]. DM employs advanced statistical tools to analyze the big data available through DWs and other sources to identify possible relationships, patterns and anomalies and discover information or knowledge for rational decision making [13, p. 590]. DW extracts or obtains its data from operational databases as well as from external open sources, providing a more comprehensive data pool including historical or current data [13, p. 590]. Big data analytics also uses statistical modelling (SM) to learn something that can support decision making [2]. Visualization techniques as an important part of big data analytics make knowledge patterns and information for decision making in a form of figure or table or multimedia. In summary, big data analytics can facilitate business decision making and realization of business objectives through analyzing current problems and future trends, creating predictive models to forecast future threats and opportunities, and analyzing/optimizing business processes based on involved historical or current data to enhance organizational performance using the mentioned techniques [12]. Therefore, big data analytics can be represented below.

$$\begin{aligned}\text{Big data analytics} = {} & \text{Big data} + \text{data analytics} + \text{DW} + \text{DM} + \text{SM} \\ & + \text{ML} + \text{Visualization} + \text{optimization}\end{aligned} \tag{1}$$

Where + can be explained as “and”. This representation reveals the fundamental relationship between big data, data analytics and big data analytics, that is, big data analytics is based on big data and data analytics, as illustrated in Fig. 1. It also shows that computer science and information technology play a dominant role in the development of big data analytics through providing sophisticated techniques and tools of DM, DW, ML and visualization [2]. SM and optimization still plays a fundamental role in the development of big data analytics, in particular in big data prescriptive analytics [11].

It should be noted that the above equation is a concise representation for the technological components of big data analytics whereas the proposed ontology of big data analytics in this Section is to look at what big data analytics constitutes at a relatively high level. We will consider the big data descriptive, predictive and prescriptive

analytics as one dimension, and the technological components of big data analytics as another dimension. Then we will provide this 2-dimension analysis as a future research work.

3 Business Intelligence and Big Data Analytics

This section will examine business intelligence (BI) and its relationships with big data analytics.

BI has drawn increasing attention in academia and business over the past two decades, although the term was already coined in 1958 by an IBM scientist [18]. There are many different definitions on BI. For example,

- BI is defined as providing decision makers with valuable information and knowledge by leveraging a variety of sources of data as well as structured and unstructured information [19].
- BI refers to as a collection of information systems (IS) and technologies that support managerial decision makers of operational control by providing information on internal and external operations [15].
- BI is a framework consisting of a set of concepts, theories, and methods to improve business decision making by using fact-based support systems [5].

The first definition of BI emphasizes information and knowledge for decision makers. The second definition stresses "a collection of ISs and technologies" while specifies the decision makers to "managerial decision makers of operational control", and information to "information on internal and external operations". The last definition emphasizes "a set of concepts, theories, and methods to improve business decision making". Based on the above analysis, BI can be defined as a set of theories, methodologies, architectures, systems and technologies that support business decision making with valuable data, information and knowledge. This definition reflects the evolution of BI and its technologies from decision support systems (DSS) and its relations with data warehouses, executive information systems [8].

The principal tools for BI include software for database query and reporting (e.g. SAP ERP, Oracle ERP, etc.), tools for multidimensional data analysis (e.g. OLAP), and data mining (e.g. predictive analysis, text mining, web mining) [20]. Data warehousing is also considered as a foundation of BI [5].

Based on the previous subsection's discussion, big data analytics can be considered a part of BI [5], because it "supports business decision making with valuable data, information and knowledge" [2]. Both BI and big data analytics are common in emphasizing either valuable data or information or knowledge. BI involves interactive visualization for data exploration and discovery, for them Tableau, QlikView and Tibco's Spotfire are BI tools for interactive visualization for data exploration and discovery [21]. These BI tools are also considered as the tools of big data analytics. This implies that BI and big data analytics share some common tools to support business decision making.

Currently, BI is based on four cutting-age technology pillars of cloud, mobile, big data and social technologies [17, 22], each of these pillars corresponds to a special kind

of web services, that is, cloud services, mobile services, big data services and social networking services; all these constitute modern web services [17]. Each of these services has been supported by analytics services and technologies [2]. They are effectively supported also by big data analytics as a service and technology, as shown in Fig. 2.

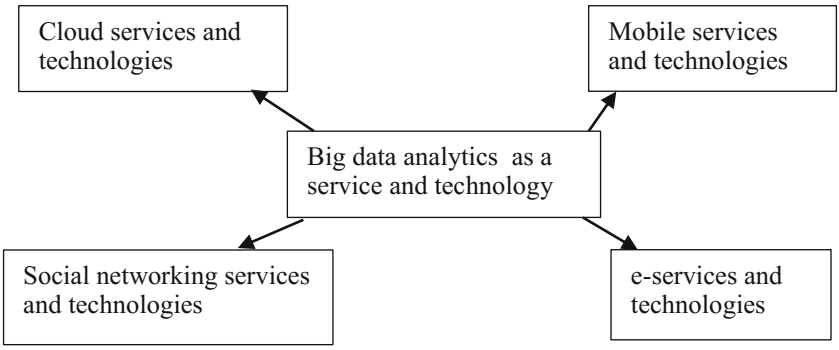


Fig. 2. Interrelationship between big data analytics and web services [2]

It should be noted that for the state-of-art web services, Sun and Yearwood [17] explores that web services mainly consist of mobile services, analytics services, cloud services, social networking services, and service as a web service. In reality, each of them involves sophisticated ICT technologies. Then, technologies are added to mobile services, analytics services, cloud services and social networking services in [2]. Here we emphasize big data analytics as a service and technology at the center to support cloud services and technologies, social networking services and technologies, mobile services and technologies, e-services and technologies to reflect the big data and analytics as an emerging new service and technology. The readers can easily find practical examples to reflect this trend. We do not go into it anymore because of limitation of space.

Based on IDC's prediction for the IT market in 2014 [22], spending on big data will explode and grow by 30 %, to $14 + billion, in which, big data analytics services will experience an explosive growth. The spending on big data analytics services will exceed $4.5 billion, growing 21 %. The number of providers of big data analytics services will triple in three years. This means that big data analytics as a service and technology has become an important emerging market, together with cloud services, mobile services and social networking services [2]. All these four services and the technologies shape the most important markets for e-commerce and e-services [17].

Furthermore, BI is a more general concept for improving business performance and business decision making. Big data analytics is a pivotal part for developing BI, at least from a technological viewpoint and data viewpoint. From a technological viewpoint, big data analytics is data-driven and business oriented technique and facilitates business decision making and then improves BI [2]. From a data viewpoint, big data analytics relies on data analytics and big data which have become a strategic natural resource for every organization, in particular for multinational organizations as well as for e-commerce and e-services. Discovering secrets from databases, data warehouses, data

marts and the Web has become the central topics for business operations, marketing and BI. This is just the task of big data analytics.

4 BASOA: Big Data Analytics Services Oriented Architecture

This section proposes a big data analytics service oriented architecture (BASOA) and then examines each of the main players in the BASOA. Different from the traditional SOA [23], the proposed BASOA specifies general services to big data analytics services, as showing in Fig. 3. We use BA in this architecture, BASOA, to represent big data analytics, which implies that big analytics (BA) can represent big data analytics briefly. This is reasonable because we have big data and big analytics, both are originally from data and analytics respectively.

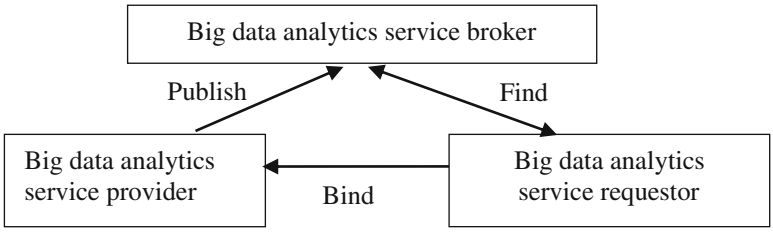


Fig. 3. BASOA: a big data analytics SOA

In this BASOA, big data analytics service provider, big data analytics service requestor, big data analytics service broker are three main players. In what follows, we will look at each of these in some detail, taking into account BI.

Big data analytics service requestors include organizations, governments and all level business decision makers such as CEO, CIO and CFO as well as managers. Big data analytics service requestors also include business information systems and e-commerce systems. Big data analytics service requestors require big data analytics services including information analytics services, knowledge analytics services, business analytics services with visualization techniques to provide knowledge patterns and information for decision making in a form of figure or table or report [24]. More generally, big data analytics service requestors include people who like to make decisions or acquire information based on analytical reports provided by big data analytics service provider [2]. Therefore, a person with smartphone receiving analytics services is also a big data analytics service requestor [12].

Big data analytics service brokers are all the entities that facilitate the development of big data analytics services, which include popular presses, traditional media and social media, consulting companies, scholars and university students, and so on [2]. All these use a variety of methods and techniques to improve the better understanding of big data analytics services in general and data analytics, business analytics, web analytics, and their services in particular [2]; all these have been offered to university students as a course material or content in business and computing areas to some extent in recent years. McKinsey Consulting (http://www.mckinsey.com/), Boston Consulting Group

(BCG), and IDC as big data analytics service brokers have played an important role in pushing big data analytics in businesses and enterprises, just as they promote "big data". Gartner and Forrester are also famous big data analytics service brokers in the world [12].

Big data analytics service providers include analytics developers, analytics vendors, analytics systems or software and other intermediaries that can provide analytics services. Recently, web analytics service (WAS) providers are important big data analytics service providers. A WAS provider, for example, Adobe Marketing Cloud (http://www.adobe.com/au/solutions/digital-marketing.html), aggregates and analyses blog data about the online behaviors of users who visited the client's website, then they evaluate a variety of analytical reports concerning the client's customer online behaviors that the client wishes to understand. This can then facilitate their strategic business decision making [25]. Application service providers (ASPs) can also provide web analytics in a hosted ASP model with quicker implementation and lower administrative costs [25]. Analytics developers provide analytic tools with extensive data extraction, analytics and reporting functionality such as Piwik, CrawlTrack [26]. Google is not only a search engine provider, but also a WAS vendor, because Google provides Google Analytics (http://www.google.com/analytics/), a big data analytics, with good tracking tools. In fact, most hosting websites, like Baidu, also provide these similar big data analytics services. A mobile phone company can provide big data analytics services to the customers with smartphone [12]. For example, Mobile App Analytics (http://www.google.com/analytics/mobile/), a part of Google Analytics, is also a mobile big data analytics services provider that helps the smartphone customers to discover new and relevant users through traffic sources reports. Mobile App Analytics plays a role of integration and gets engaged through event tracking and flow visualization, and sets and tracks the goal conversions one wants most: purchases, clicks, or simply time spent on the app. More generally, many information systems have contained an analytics app as a system component to generate table, diagram or report. All these kinds of information systems can be considered as big data analytics service providers. The big data analytics services providers on the Web include Amazon, Google and Microsoft [2].

5 Applying BASOA to BI

This section looks at how to apply the proposed BASOA to enhance BI in some detail.

Analytics as a service (AaaS) is a relatively new concept that has emerged as a rapidly growing business sector of web analytics industry, which provides efficient web log analytic services for firm-level customers [17]. BAaaS (Big data analytics as a service), as discussed in the BASOA above, means that an individual or organization or information system uses a wide range of analytic tools or apps wherever they may be located [12]. BAaaS has the ability to turn a general analytic platform into a shared utility for an enterprise with visualized analytic services [12]. An analytics service can be available on the Web or used by smartphone. Therefore, big data analytics services include e-analytics services or web analytics services (WAS) [2]. BAaaS is gaining popularity rapidly in business, e-commerce, e-service, and management in recent years. For example, BAaaS model has been adopted by many famous web companies such as

Amazon, Microsoft, and eBay [12]. The key reason behind it is that the traditional hub-and-spoke architectures cannot satisfy the demands driven by increasingly complex business analytics [12]. BAaaS promises to provide decision makers with visualizing much needed big data. Cloud analytics is an emerging alternative solution for big data analytics [2].

As previously defined, BI is "a set of theories, methodologies, architectures, systems and technologies that support business decision making with valuable data, information and knowledge". BASOA is an architecture for supporting business decision making with big data analytics services. The theory of big data analytics providers, brokers and requestors of the BASOA can facilitate the understanding and development of BI and business decision making. For example, from a deep analysis of the BASOA, an enterprise and its CEO can know who are the best big data analytics providers and brokers in order to improve his business, market performance, and competition.

We surveyed 71 information technology managers at the Association for Education in Journalism and Mass Communication (AEJMC) in Montreal during August 6–9, 2014 [2], to collect data concerning the enterprise-level acceptability of the BASOA concept. These results indicate some preliminary support for the BASOA concept of having service brokers work with service requesters and providers similar to the way private mortgage and loans work in the USA. Based on this preliminary enterprise acceptability of this BASOA model, we propose that more research be done to investigate how it could be used.

6 Related Work and Discussion

We have mentioned a number of scholarly researches on data analytics, big data analytics, and BI. In what follows, we will focus on related work and discussion on ontology of big data analytics, and the work of SAP as well as incorporation of big data analytics into BI.

Why does big data analytics really matter for modern business organizations? There are many different answers to this question from different researchers. For example, Davis considers that the current big data analytics has embodied the state of art current development of modern computing [27], which has been reflected in Sect. 2. Gandomi and Harder [7] discuss how big data analytics has captured the attention of business and government leaders through decomposing big data analytics into text analytics, audio analytics, video analytics, social media analytics, and predictive analytics. This implies that data has been classified into text data, audio data, video data, and social media data in [7].

Big data analytics and BI have drawn an increasing attention in the computing, business, and e-commerce community. For example, Lim et al. [5] examine business intelligence and analytics by focusing on its research directions. They consider business intelligence and analytics (BIA) as a current form replacing the traditional BI, whereas we still consider BI and big data analytics are two different concepts, although they have close relationships and share some commons. Fan et al. [6] provide a marketing mix framework for big data management through identifying the big data sources, methods, and applications for each of the marketing mix, consisting of people, product, place,

price and promotion. However, what is the relationship between marketing intelligence and BI in terms of big data analytics should have been mentioned in their work [6].

Ontology has been important in computer science and artificial intelligence [16]. A basic search in Google scholar (i.e. article title and key words) reveals that there are few publications entitled "ontology of big data analytics". We then explored it and put it as a part of this research through updating our early work on data analytics, business analytics and big data analytics [2]. We explore the interrelationship among big data analytics, big data descriptive analytics, big data predictive analytics, and big data prescriptive analytics using the proposed ontology. The results reported in this paper on ontology of big data analytics and big data analytics equation are an extension and development of our early work [2] by adding optimization and ML to the equation. This is only a beginning for providing a relatively comprehensive ontology of big data analytics. In this direction, we will investigate more academic reviewed sources as a future work to develop an ontology of big data analytics with three levels for each related analytics: big data, methods and applications based on the method of Fan et al. [6]. Such an investigation would become an important guide for the research and development of big data analytics and BI.

SAP, one of the leading vendors of ERP [28], has introduced its enterprise service-oriented architecture [20, p. 383]. SAP's architecture specifies general services to enterprise services whereas our BASOA model specifies general services to big data analytics services. Big data analytics services should be a part of state-of-the-art e-commerce services [17], and then the proposed BASOA can be considered as a concrete application for the enterprise service-oriented architecture of SAP. However, SAP's enterprise systems focus on key applications in finance, logistics, procurement and human resources management as an ERP system. We conceive that our BASOA will be incorporated into the next generation enterprise systems integrating SCM, CRM, and KM systems, and e-commerce systems. This is also the motivation of our proposed BASOA.

7 Conclusion

This paper proposed an ontology of big data analytics, and looked at the relationship between big data analytics and BI. This paper also presented a big data analytics service oriented architecture (BASOA) and discussed how to use BASOA to enhance BI. The preliminary analysis on the collected data shows that this proposed BASOA is viable for facilitating the development of BI. The proposed approach in this paper might facilitate research and development of big data analytics, business analytics, BI, e-commerce, and e-services.

In the future work, besides mentioned in the previous sections, we will analyse the foregoing collected data vigorously and explore enterprise and e-commerce acceptability of BASOA for BI. We will also explore big data analytics and its applications in e-commerce and cloud services, and realize BASOA using intelligent agents technology [29], where we will also look at some implementation related issues such as how to collect, store, and process big data – by whom, for what, access rights, and many more.

References

1. Chen, C.P., Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf. Sci. **275**, 314–347 (2014)
2. Sun, Z., Strang, K., Yearwood, J.: Analytics service oriented architecture for enterprise information systems. CONFENIS 2014, Hanoi, 4–6 December 2014. In: Proceedings of iiWAS2014, pp. 506-518. ACM Press (2014)
3. Sun, Z., Firmin, S., Yearwood, J.: Integrating online social networking with e-commerce based on CBR. In: Proceedings of the 23rd ACIS 2012, Geelong, 3–5 December 2012
4. Vesset, D., McDonough, B., Schubmehl, D., Wardley, M.: Worldwide Business Analytics Software 2013–2017 Forecast and 2012 Vendor Shares (Doc # 241689), 6 2013. [Online]. http://www.idc.com/getdoc.jsp?containerId=241689. Accessed 28 June 2014
5. Lim, E., Chen, H., Chen, G.: Business intelligence and analytics: research directions. ACM Trans. Manage. Inf. Syst., vol. 3, no. 4, Article 17, pp. 1–10 (2013)
6. Fan, S., Lau, R.Y., Zhao, J.L.: Demystifying big data analytics for business intelligence through the lens of marketing mix. Big Data Res. **2**, 28–32 (2015)
7. Gandomi, A., Haider, M.: Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manag. **35**, 137–144 (2015)
8. Holsapplea, C., Lee-Postb, A., Pakath, R.: A unified foundation for business analytics. Decis. Support Syst. **64**, 130–141 (2014)
9. van der Meulen, R., Rivera, J.: Gartner Says Worldwide Business Intelligence and Analytics Software Market Grew 8 Percent in 2013, 29 4 2014. http://www.gartner.com/newsroom/id/2723717. Accessed 28 June 2014
10. Beal, V.: Big data analytics 2014. http://www.webopedia.com/TERM/B/big_data_analytics.html. Accessed 20 June 2014
11. Minelli, M., Chambers, M., Dhiraj, A.: Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Chinese Edition 2014 edn. Wiley & Sons, Hoboken (2013)
12. Delena, D., Demirkanb, H.: Data, information and analytics as services. Decis. Support Syst. **55**(1), 359–363 (2013)
13. Coronel, C., Morris, S., Rob, P.: Database Systems: Design, Implementation, and Management, 11th edn. Course Technology, Cengage Learning, Boston (2015)
14. Cramer, C.: How Descriptive Analytics Are Changing Marketing, 19 May 2014. http://www.miprofs.com/wp/descriptive-analytics-changing-marketing/. Accessed 6 July 2015
15. Turban, E., Volonino, L.: Information Technology for Management: Improving Strategic and Operational Performance, 8th edn. Wiley, Danvers (2011)
16. Gruber, T.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. **43**(5–6), 907–928 (1995)
17. Sun, Z., Yearwood, J.: A theoretical foundation of demand-driven web services. In: Demand-Driven Web Services: Theory, Technologies, and Applications, IGI-Global 2014, pp. 1–25 (2014)
18. Luhn, H.P.: A Business Intelligence System. IBM J. Res. Dev. **2**(4), 314–319 (1958)
19. Sabherwal, R., Becerra-Fernandez, I.: Business Intelligence: Practices, Technologies, and Management. John Wiley & Sons Inc, Hoboken (2011)
20. Laudon, K.G., Laudon, K.C.: Management Information Systems: Managing the Digital Firm, 12th edn. Pearson, Harlow (2012)
21. Brust, A.: Gartner releases 2013 BI Magic Quadrant 2013. http://www.zdnet.com/gartner-releases-2013-bi-magic-quadrant-7000011264/. Accessed 14 February 2014

22. IDC: IDC Predictions 2014: Battles for Dominance — and Survival — on the 3rd Platform, December 2013. http://www.idc.com/getdoc.jsp?containerId=244606. Accessed 13 February 2014
23. Papazoglou, M.P.: Web Services: Principles and Technology. Pearson Prentice Hall, Harlow (2008)
24. Kauffman, R.J., Srivastava, J., Vayghan, J.: Business and data analytics: new innovations for the management of e-commerce. Electron. Commer. Res. Appl. **11**, 85–88 (2012)
25. Park, J., Kim, J., Koh, J.: Determinants of continuous usage intention in web analytics services. Electron. Commer. Res. Appl. **9**(1), 61–72 (2010)
26. Laudon, K., Laudon, J.: Management Information Systems-Managing the Dgital Firm. Person, Boston (2012)
27. Davis, C.K.: Viewpoint beyond data and analytics- why business analytics and big data really matter for modern business organizations. CACM **57**(8), 39–41 (2014)
28. Elragal, A.: ERP and Big Data: The Inept Couple. Procedia Technol. **16**, 242–249 (2014)
29. Sun, Z., Finnie, G.: Intelligent Techniques in E-Commerce: A Case-Based Reasoning Perspective, p. 2010. Springer, Heidelberg (2004)

Linked Relations Architecture for Production and Consumption of Linksets in Open Government Data

Petar Milić(✉), Nataša Veljković, and Leonid Stoimenov

Faculty of Electronic Engineering, University of Niš,
Aleksandra Medvedeva 14, Niš, Serbia
milicpetar86@gmail.com,
{natasa.veljkovic,leonid.stoimenov}@elfak.ni.ac.rs

Abstract. Linking open data in government domain, can lead to creation of new services and information as well as discovery of new ways to perform queries and get results in accessible, machine processable and structured manner. To reach the full potential of open government data more relations between data should be discovered. The interconnection of open government data and semantic description of their relations can bring new aspect of producing and consuming the data. In this paper we investigate issues for producing and utilizing open government data with special focus on dataset relations. We have proposed the Linked Relations (LIRE) architecture for relations creation between datasets and a basic RDF model of relation between two datasets. The architecture contains different modules that perform analysis of datasets attributes and suggest the type of relation between the datasets. It can be utilized by open data portals for creating relations between datasets belonging to different public agencies and government sectors. An idea presented in this paper is made available as CKAN plugin.

Keywords: Open government data · Linked data · Linkset · Dataset relations · CKAN open data portal

1 Introduction

Publication of open government data (OGD) leads to more openness, transparency and efficiency of public administration. It also brings benefits for citizens by influencing development of government services for society, hence producing better public service outcomes. Open data philosophy suggests that data should be published in open formats and in ways that make them accessible, readily, reusable and available to the public, business and government sector [1]. Following this approach, data can be easily consumed by both web developers and common users. Having in mind that most of the published data comes in original (raw) format, beneficiary contribution is not negligible, we can even say that it is immense. Every reuse of data adds new value and creates new knowledge enabling data lifecycle to expand and evolve.

Web of Data represents decentralized and heterogeneous sources of information interlinked through typed links. This is achieved by publishing structured data in

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 212–222, 2015.
DOI: 10.1007/978-3-319-25013-7_17

Resource Description Format (RDF) using URIs. RDF is the format on which is based Semantic Web, as its use of URIs allows data to be identified by reference and linked with other relevant data by subject, predicate or object. Making OGD available in the Web of data, makes them publicly available, and practically expand Web of data space, allowing their discovery and usage. Kalampokis et al. [2] claim that real value of OGD is revealed with linking data which provides unexpected and unexplored insights into different domains and problem areas. Linked Government Data (LGD) are actually OGD that runs on Semantic Web's kerosene – metadata. Metadata provide documentation, context and necessary background information.

According to Sheridan and Tennison [3] the Semantic Web standards are mature and powerful, but there is still a lack of practical approaches and patterns for publishing OGD. Nevertheless, over the last few years there is increasing number of governments that are publishing their OGD data as linked data. The adoption of the LGD has led to the extension of the government open data space, connecting data from diverse domains such as economy, finance, medicine, statistics and others to enable new types of applications. With LGD paradigm users can browse one data source and then navigate along links to related data sources. Most promising implementation of LGD is based on Semantic Web philosophy and technologies. Tim Berners-Lee [4] noted this and he gave instructions how to link data and to include government data into the Web of Data.

By following the Open Government movement [5] in the world, many governments have published their open data through the open data platforms [6]. One of the most utilized open-source solutions for publishing open datasets is the CKAN's (Comprehensive Knowledge Archive Network) open data platform [7]. This platform enables both back-end and front-end interface, used respectively for publishing/modifying and searching/reviewing open datasets. What this platform doesn't offer is the possibility to link datasets between each-other, creation of meaningful relations between them and publishing this data as linked data. This tackled our minds and we wanted to generalize this problem and create a common architecture that could be applied to other open data platforms as well and to suggest possible relations between datasets and enable their linking.

In this paper we explore area of dataset relations for producing and utilizing linked government data. We describe architecture that determines type of relation between datasets, in order to take advantage of the relations for dataset linking. The architecture contributes to defining a model of semantic representation of dataset relations and their automatic production. This leads towards production of quality LGD in line with their interlinking and integration. The remaining of the paper is organized as follows. In Sect. 2 we review related work. In Sect. 3 we propose an architecture for modelling and linking dataset relations. Section 4 presents the visual tool for creating relations between datasets, developed on the basis of architecture described in Sect. 3 that shows benefits of its use. Finally, in Sect. 5 conclusions are drawn along with the future work.

2 Related Work

A group of authors [8] proposed architecture for integrating Public Sector Information (PSI) catalogs via the activities and components essential for discovery, allowing the presentation of catalogs in standardized form, facilitating search and retrieval across

resources. This architecture requires downloading and transforming catalogs with retrievable records into a common schema language format along with addressing semantic heterogeneity with schema matching and statistical analysis of ontology structures. Pioneers in LGD, UK and USA have shown that creating high quality Linked Data from raw data files requires considerable investment into reverse-engineering, documenting data elements, data cleanup, schema mapping and instance matching [9].

If we want that government datasets become linked government datasets, semantics must be added to them. Appropriate rules tell us how to describe and how to establish links between them. For linking datasets, Alexander et al. [10] propose Vocabulary of Interlinked Datasets (VOID), a vocabulary that allows to formally describe linked RDF datasets. It defines classic LOD and 3rd-party case. In the LOD case the linkset is a subset of one of the two involved datasets, while in 3rd-party case a third dataset is involved that actually contains the linkset.

Interoperability between government datasets and bringing them closer to the Web of Data is also discussed in [11]. The authors designed DCAT vocabulary, based on exploration of seven existing open data portals to allow expression of datasets in the RDF data model. They conduct feasibility study to prove their claim that different catalogues can be rendered in proposed vocabulary.

Many authors in literature deals with proposing systems for production and consumption of LGD. Ding et al. [12] suggest the use of LGD ecosystem for LGD data production and consumption as a Linked Data – based system where users manage and consume open government data in connection with online tools, services and societies. It supports large-scale LGD production, promote LGD consumption and grow of the LGD community. TWC LGD ecosystem from [12] is based on converting raw OGD datasets into linked data and their integration with other resources. Kalampokis et al. [13] give classification scheme for OGD, where they showed technological and organizational approach for provision of linked data based on relevant literature. The proposed architecture links decentralized data with maintaining a list of available resources in the area and assigning a URI to each of them. This solution is intended to use in single government cases to link data in different datasets belonging to different public agencies and government sectors.

Schmachtenberg et al. [14] present an overview of the linkage relationships between datasets in the form of an updated LOD cloud diagram based on data that can actually be retrieved by a Linked Data crawler. They consider that two datasets can be linked if there exists at least one RDF link between resources belonging to the datasets. Using metadata with appropriate metadata architecture can bring benefits for LGD publication and use, along with improving the ability for finding and interpreting of LGD data, creating order within datasets, comparing, correct interpretation, accessibility, visualization and other benefits, as discussed in [15].

Janssen, Estevez and Janowski state in [16] that successfully linking of datasets requires understanding the data's context sensitive meaning. They claim that collected data from organizations which do not always anticipate its full potential use, might not sufficiently align with other datasets or possible relationships among datasets are unknown.

To the best of our knowledge, there are not work that deals with automatic production of linked datasets. Some of them [8, 9, 12] requires intervention of users which is a time-consuming process and does not specifies rules for their interlinking. Kalampokis [13] and Schmachtenberg [14] discuss on linking datasets based on their semantic description without going deeper in the OGD datasets. This work does not tackle informations hidden in published OGD datasets, which by our opinion can help in linking OGD datasets. Zuiderwijk et al. [15] claims that metadata can give potential in realization of all benefits from linked data that exists in literature. Similarly, Janssen et al. [16] points to the fact that the relations between datasets are unknown and that they can contribute to better linking of datasets.

Based on the approaches for linking OGD datasets mentioned in previous paragraphs, we got the idea to explore relations between datasets to check whether they can be used to produce linked datasets. According to that, we have developed an architecture for visual creation of relations between datasets, and their automatic linking. The so called LIRE (Linked Relations) architecture was created at first to enrich CKAN open data portal with a tool that enables automatic creation and deletion of linksets. In the following section we will explain more thoroughly the proposed architecture.

3 Linked Relations Architecture – LIRE

LIRE system architecture, outlined in Fig. 1, consists of different interconnected components, each with specially assigned tasks. Implemented functionalities are available through a single workbench. LIRE enables users to find, manage, integrate, publish and reuse relations between datasets. It promotes production and consumption of linksets, semantic data that describe relations between datasets.

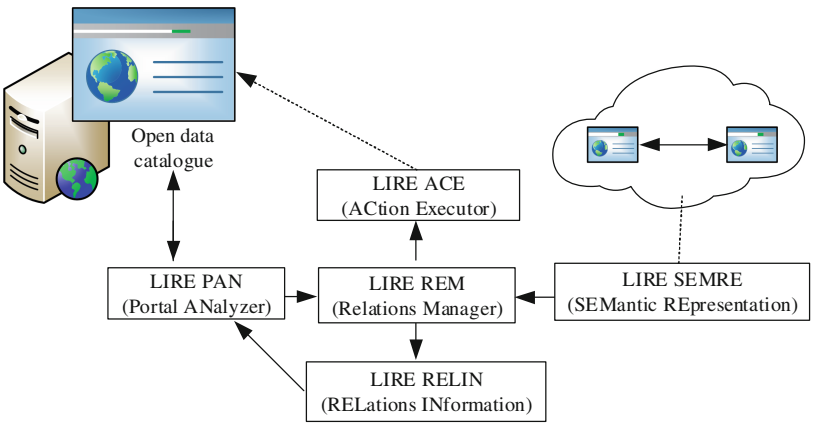


Fig. 1. LIRE system architecture

LIRE consists of the following modules:

- Portal Analyzer – PAN. This module is the entry point of LIRE. PAN enables filtering of datasets on different parameters (by tag, organization, group, all datasets or random datasets). It prepares datasets for processing for REM module and gives necessary information about datasets to RELIN module for their visual display.
- Relations Manager – REM. Creating, managing and determining type of relations between datasets based on their metadata are carried out in REM. This module implements a model on which is based determination of the type of relation, described in detail in the next section. REM examines datasets metadata to determine similarity between them for possible relation creation. Module does not limit the user in choosing the type of relations, so that the user does not have to apply the type of relation that is suggested by REM, but can apply selected, if it differs from the proposed one. It calls ACE module to execute pending actions to the OGD portal and RELIN module to refresh display after editing.
- Action Executor ACE – This module executes actions created by REM to store results of editing in the OGD platform. Supported actions are "CREATE" and "DELETE".
- Relations Information – RELIN is module used for visualization of relations. It enables short preview of information of datasets in REM module to enable user to decide whether to relate datasets or not, and user interface for managing datasets relations. Also incorporates jQuery and CSS libraries for visualizing datasets relations and their graphical management. Every dataset is represented with graphical element that contains information on dataset's description, tags, formats and existing relations.
- Semantic Representation – SEMRE module creates semantic representation of any existing relations. This semantics is created based on the model of RDF graph, described in detail in Sect. 3.2. Implemented RDF graph model is based on voID (vocabulary of Interlinked Datasets) vocabulary, because of its simplicity for describing linked datasets.

3.1 Creating Relations with LIRE

Relations Manager deals with managing of relations between datasets. It examines data that describe datasets to determine type of relation. After examination, determined type is suggested to user who can accept suggested solution or choose another one. Using developed relation suggestion models, described in Table 1, that are based on presence/absence of selected datasets metadata, REM module of LIRE architecture can determine whether datasets can have one of the following relations: parent_of, child_of, links_from, links_to. The child_of relation model consists of thirteen conditions, listed as C1-C13, where each condition examines certain dataset property or combination of properties on a true/false basis. If all conditions are met then relation between two datasets is of type child_of. The conditions for child_of relation can be used also for determining whether the relation between two dataset is of type parent_of, but with following modifications: conditions C4, C5, C7 and C8 should be less than, while C10-C12 should be greater than.

Table 1. Models for relations child_of and links_from

CHILD_OF		LINKS_FROM	
C1. Number of same/similar tags between two datasets	>0	C1. Number of same/similar tags between two datasets	>0
C2. Do they belong to the same organization	true	C2. Whether they are open	true
C3. Do they belong to the same group	true	C3. Whether the number of the same/similar resource formats of the first dataset is greater than the number of the same/similar resource formats in the second dataset	>
C4. Whether the number of the same/similar tags of the first dataset is greater than the number of the same/similar tags in the second dataset organization	>	C4. Whether the five star index of the both datasets is higher than 3	>3
C5. Whether the number of the same/similar tags of the first dataset is greater than the number of the same/similar tags in the second dataset group	>	C5. Whether they have at least one linked format in its resources	true
C6. Are they linked via links in extra field	true	C6. Whether they have at least one machine processable format	true
C7. Whether the number of the same/similar resource formats of the first dataset is greater than the number of the same/similar resource formats in the second dataset	>	C7. Whether the first dataset was created before the second	<
C8. Whether the first dataset was created after the second	>	C8. Whether the descriptions of two datasets are similar	>n
C9. Whether the descriptions of two datasets are similar	>n		
C10. Whether the number of total views of the first dataset is less than the number of total views of the second dataset	<		
C11. Whether the number of recent views of the first dataset is less than the number of recent views of the second dataset	<		
C12. Whether the five star index of the first dataset is less than the five star index of the second dataset	<		
C13. Whether they are open	true		

For links_from relation there are eight conditions, listed as C1-C8. If all conditions are met then relation between two datasets is of type links_from. These conditions can be used for the determining whether the relation between two datasets is of type links_to with following modifications: condition C3 should be less than and C7 greater than.

3.2 Creating Semantics of Relations with LIRE

Modelling relations between OGD datasets using linked data principles and techniques can add more semantics to government data, enabling thus easier search and retrieval of information by using semantic tools. Adding semantics to OGD is achieved through RDF description of datasets with help of Dublin Core and DCAT vocabularies [17, 18]. Dublin Core expresses metadata that describe dataset in RDF for direct machine processing through most well-known and basic terms, while DCAT facilitates interoperability and increases discoverability for easy consume of LGD.

LIRE architecture has SEMRE component which carries out modelling of relations with use of voID (vocabulary of Interlinked Datasets) vocabulary, because voID is one of the most widespread vocabularies for LGD and has a feature called linkset. A linkset is a collection of RDF links where an RDF triple has subject and object described in different datasets [19]. Vocabulary voID is convenient for use in our case because it is naturally intended for describing linked datasets. Knowing linkset structure, we can define a basic RDF model of relation between two OGD datasets implemented in SEMRE (Fig. 2).

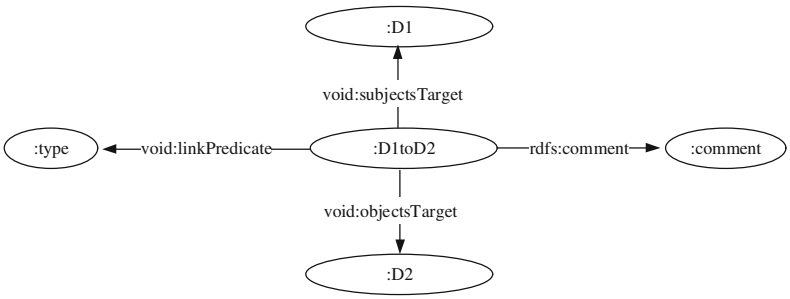


Fig. 2. RDF graph of relation between two datasets implemented in LIRE SEMRE

Implemented RDF graph model can be described using two semantic web data formats Turtle and RDF + XML. In Turtle, it would be:

```
@prefix void: <http://rdfs.org/ns/void#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
:D1toD2 a void:Linkset;
void:subjectsTarget :D1;
void:objectstarget :D2;
void:linkPredicate :type;
```

```
rdfs:comment :comment;
```

Represented with RDF+XML, it would look like:

```
<rdf:RDF>
<void:linkSet>
<void:subjectsTarget>D1</void:subjectsTarget>
<void:objectsTarget>D2</void:objectsTarget>
<void:linkPredicate>type</void:linkPredicate>
<rdfs:comment>comment</rdfs:comment>
</void:linkSet>
</rdf:RDF>
```

In Turtle code D1toD2 represents the name of linkset, i.e. the name of the relation, and it is identified by void:linkset statement. It is also a part of URI of appropriate RDF/XML syntax. Terms D1 and D2 are dataset's names represented by void:subjectsTarget and void:objectsTarget statements respectively. In both data formats type is the type of relation (parent_of, child_of, etc.) represented by void:linkPredicate. Relation description is contained in comment element identified by rdfs:comment.

The appropriate mapping between dataset relation elements and voID vocabulary is illustrated on Fig. 3.

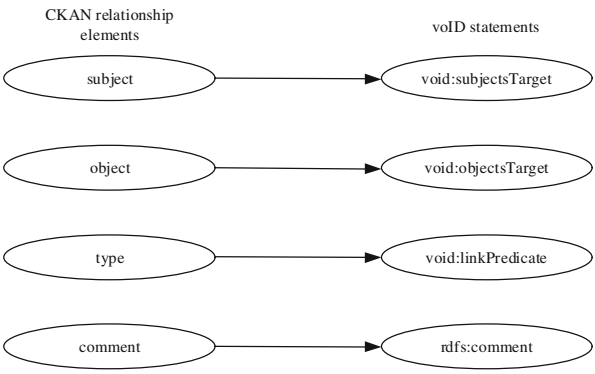


Fig. 3. Mapping between CKAN relations properties and voID statements conducted in SEMRE

SEMRE uses voID description of dataset relations with aim to offer users more data related with it and to enable easy access, search and retrieval of information. In this manner SEMRE enables access to LGD from semantic web applications. It also offers a mechanism to implement a semantic description of dataset relations into the open data catalog.

The RELIN component of architecture enables user interface for creating related datasets. Every dataset is represented with graphical element that contains information on dataset's description, tags, formats and existing relations.

4 LIRE Architecture Deployment for CKAN Data Catalogue

To demonstrate the value of LIRE architecture, we have deployed the architecture as plugin for CKAN open data portal. The plugin is in beta phase now, but it will be soon available from the CKAN's online plugin repository. To present the use case of using LIRE as plugin we have installed CKAN platform on local computer, and uploaded to it few datasets from datahub.io. To see existing relations between datasets, user filters the display by using one of the following parameters: datasets per tag, per group, per organization or random number of datasets. If he skips filtering, all datasets from portal and their relations will be loaded. RELIN component gives output depicted on Fig. 4. All datasets matching user filters and their related datasets are presented in the page. With the given datasets, user can: create, update and delete relations between the datasets and to select datasets for which want to obtain a semantic description of the relations in selected format.

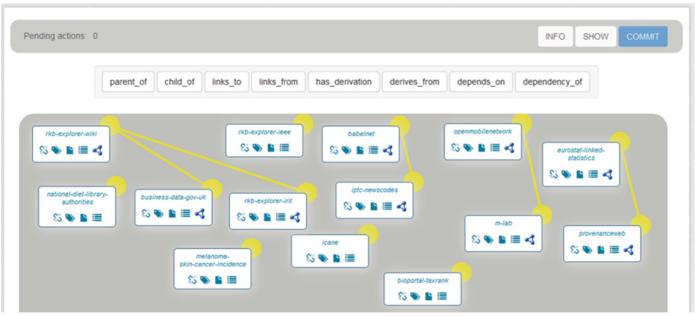


Fig. 4. Visual representation of dataset relations

User can perform following operations: remove or create relation between datasets. Every user action is saved in the list of actions pending to be executed in CKAN. After user finishes changes on the datasets, he needs to commit actions to CKAN. Removing existing relation is enabled through the user interface by clicking on the relation and choosing option delete from the context menu. When a user chooses to create relation between two datasets, for example *rkb-explorer-ieee* and *rkb-explorer-wiki* with child_of relation, firstly he needs to choose relation type from the application menu. Direction of the connection determines subject and object of the dataset relation, starting dataset is subject and ending dataset is object. Application examines meta-data of these datasets based on model described in Table 1 and returns results to the user. In these concrete example, user pick was child_of, and this type is not matched with the one proposed by the model links_from (Fig. 5). User can choose to proceed with his action and relate datasets in child_of relation or take the suggestion from LIRE and go for links_from relation.

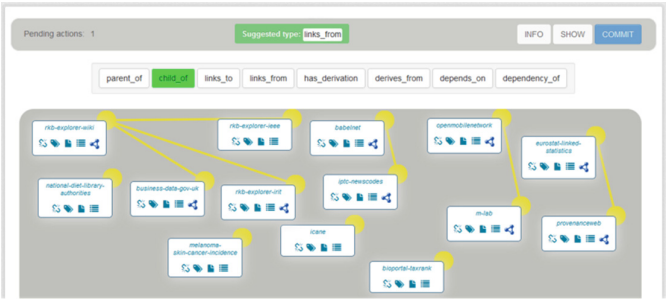


Fig. 5. Relating datasets

5 Conclusion

In this paper we presented an architecture for relating datasets and modelling or managing their relations with linked data principles and techniques along with model for their semantic representation. It reduces effort needed to preview datasets in order to relate them based on characteristics and data that describe them. As we have shown in Sect. 2, so far there is no research in the area of datasets relations and their modelling by using linked data. Dataset relations offer local level of relating, but if we describe them by linked data, dataset can be enriched with new data and information and with added semantics. For that purpose an RDF graph model for describing relations between two datasets is defined using voID vocabulary. Mapping between CKAN relations properties and voID statements shows that there are simple way for producing linked datasets. The modular nature of the proposed architecture makes it applicable to other portals except CKAN, but it requires additional time to review and analyze the data that describe the datasets, which are an essential element of our architecture. Future work includes investigation of the possibility for the development of the model and tool for semantic management of datasets and their relations and platform for accessing linked datasets based on defined RDF model. Also, incorporating proposed linked datasets model into the CKAN will be of great help to the users and developers in creating semantic applications.

References

1. Ayers, D.: Evolving the link. IEEE Internet Comput. **11**(3), 95–96 (2007)
2. Kalampokis, E., Tambouris, E., Tarabanis, K.: Linked open government data analytics. In: Wimmer, M.A., Janssen, M., Scholl, H.J. (eds.) EGOV 2013. LNCS, vol. 8074, pp. 99–110. Springer, Heidelberg (2013)
3. Sheridan, J., Tennison, J.: Linking UK government data. In Proceedings of the WWW 2010 Workshop on Linked Data on the Web (2010)
4. Berners-Lee, T.: Linked data. http://www.w3.org/DesignIssues/LinkedData.html
5. Veljković, N., Bogdanović-Dinić, S., Stoimenov, L.: Benchmarking open government: an open data perspective. Gov. Inf. Q. **31**(2), 278–290 (2014). Elsevier

6. Veljković, N., Bogdanović-Dinić, S., Stoimenov, L.: Exploring collaboration between public administrations through the notion of open data (ICIST 2015), Kopaonik, 8–11 March, vol. 1, pp. 122–127 (2015)
7. CKAN Open data portal. http://ckan.org/
8. Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W.: Linked open government data: lessons from data.gov.uk. IEEE Intell. Syst. **27**(3), 16–24 (2012)
9. Government, H.M.: Putting the frontline first: smarter government. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228889/7753.pdf
10. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing linked datasets. In: Proceedings of the 2nd Workshop on Linked Data on the Web (2009)
11. Maali, F., Cyganiak, R., Peristeras, V.: Enabling interoperability of government data catalogues. In: Wimmer, M.A., Chappelet, J.-L., Janssen, M., Scholl, H.J. (eds.) EGOV 2010. LNCS, vol. 6228, pp. 339–350. Springer, Heidelberg (2010)
12. Ding, L., Lebo, T., Erickson, S.J., DiFranzo, D., Williams, T.G., Li, X., Michaelis, J., Graves, A., Zheng, G.J., Shangguan, Z., Flores, J., McGuinness, L.D., Hendler, A.J.: TWC LOGD: a portal for linked open government data ecosystems. Web Semant. Sci. Serv. Agents World Wide Web **9**(3), 325–333 (2010)
13. Kalampokis, E., Tambouris, E., Tarabanis, K.: A classification scheme for open government data: towards linking decentralised data. Int. J. Web Eng. Technol. **6**(3), 266–285 (2011)
14. Schmachtenberg, M., Bizer, C., Paulheim, H.: Adoption of the linked data best practices in different topical domains. In: Mika, P., et al. (eds.) ISWC 2014, Part I. LNCS, vol. 8796, pp. 245–260. Springer, Heidelberg (2014)
15. Zuiderwijk, A., Jeffery, K., Janssen, M.F.: The potential of metadata for linked open data and its value for users and publishers. JeDEM-e-J. e-Democracy Open Gov. **4**(2), 222–244 (2012)
16. Janssen, M., Estevez, E., Janowski, T.: Interoperability in big, open, and linked data-organizational maturity, capabilities, and data portfolios. Computer **47**(10), 44–49 (2014)
17. Dublin Core Metadata Initiative: DCMI Metadata Terms. http://dublincore.org/documents/dcmi-terms/
18. W3C: Data Catalog Vocabulary (DCAT). http://www.w3.org/TR/vocab-dcat/
19. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Linkset. http://www.w3.org/TR/void/#linkset

Budgetary Data (in an Open Format) Benefits, Advantages, Obstacles and Inhibitory Factors in the View of the Intermediaries of this System: A Study in Latin American Countries

Gisele da Silva Craveiro[1(✉)] and Cláudio Sonáglio Albano[2]

[1] Universidade de São Paulo (USP), São Paulo, Brazil
giselesc@usp.br
[2] Universidade Federal do Pampa (UNIPAMPA), Bagé, Brazil
claudio.albano@unipampa.edu.br

Abstract. Governments are under pressure to meet new social demands and seek new forms of management. As well as this, the fact that they make considerable use of information technology, has led to the growth of their databases, and made governments and the respective government organizations, fertile ground for open data initiatives. Having access to budgetary data and being able to make use of it (with regard to revenue and expenditure), has traditionally always aroused great interest in society. The purpose of this study is to determine the potential benefits and possible obstacles that can affect the intermediaries who take action on the basis of budgetary data. A number of intermediaries (members of society and government policymakers) from four Latin-American countries were interviewed. The results show that the structural barriers of governments have an adverse effect on their activities because they impair the quality of the information that is made available. Nonetheless, the benefits that allow a greater degree of transparency as well as the ability to reveal more knowledge of the inside operations of governments, encourage the intermediaries to take part in this "ecosystem".

Keywords: Benefits · Obstacles · Open government data · Intermediaries

1 Introduction

Public organizations (in this study regarded as governments) are being confronted with new demands on the part of society. These demands are linked to the need for greater transparency in the management of public funds, and a greater control over the services provided, as well as a greater degree of accountability on the part of their managers. In this climate, information and communication technology (ICT) is beginning to play a key role.

The Internet can be cited as one of the most important resources for this technology. The set of factors outlined here – a greater demand by society for public organizations, a widespread use of information technology and the growing use and importance of the Internet – have made feasible the rise of a platform called open government. In the

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 223–235, 2015.
DOI: 10.1007/978-3-319-25013-7_18

view of [6], under the aegis of open government, public organizations are seeking to meet the requirements of society by offering information as a means of finding a solution to some of the problems.

Among an array of opportunities created by open government initiatives, there is the prospect of being able to use open data (OD) which according to [3] means making information available on the Internet in a way that can allow it to be recycled for third parties. According to [12], open government data (OGD) involve the publication and dissemination of information of the public sector (i.e. governments) in the Web, which is shared in a rational and comprehensible format in a way that allows its reuse in digital applications or in other words, makes it legible for machines.

Following the rise of concepts (or technological platforms) of open data and open government, new players are emerging in the scene who may require new ways of using and appropriating the information made available, as well as reshaping traditional requirements. Thus the OGD projects must form a network between the government and society. It is only in this way that the assurances about the prospects of using open data effectively can be fulfilled, which means it is necessary to create and maintain an environment where the different players can interact [11]. Thus, governments and society must create an environment which benefits both and allows these sectors to grow and become more involved. Several authors claim that there are potential benefits that can be derived from open government data: economic growth and an improvement in the provision of public services among other factors.

From what has been outlined so far, undertaking this work is justified in so far as it seeks to achieve the following general objective: to find out the advantages, benefits, obstacles and inhibitory facts in the view of the OG intermediaries, who are involved in the ecosystem of budgetary data in Latin American countries. In attaining this goal, it is hoped that an answer can be found to the following research question: by having a knowledge of the expectations of the OGD intermediaries, can governments be granted better conditions to carry out projects of open government data in the budgetary "ecosystem"?

By undertaking this study it is hoped that a contribution can be made to the background of open government data by obtaining information about how the intermediaries take action, what benefits are sought that can overcome obstacles and other factors that pervade the activities carried out in the budgetary "ecosystem" of open data. Another key issue is the context of the work since it focuses on representative countries from the Latin American continent, namely Argentina., Brazil, Mexico and Uruguay.

The exposed here, we consider important to clarify that intermediaries terms and ecosystem in this work. Intermediate are players (individually or representative of governments and civil society organizations), which operate with public data, released in an open format. Ecosystem as a group of organizations and individuals (government and society), as well as structures (example - public authorities and their agencies, non-profit society organizations), technological tools (e.g. - software), legal components (example - laws) and values of society that interact in a particular government sector and society by making use of public data released in an open format.

2 Theoretical Framework

The theories and models which act as a frame of reference, are employed for this study. This entails defining concepts about open data, its eco-system and the performance of intermediaries. The framework is outlined in the following sections together with its advantages, benefits, obstacles and inhibitory factors in an OGD ecosystem which supports the data collection data and analysis.

2.1 Open Data, Its Ecosystem and Intermediaries

The advent of the platform (or concept) of open data originates in a historical process, where public organizations were always great users of information technology. This fact allied to new requirements by society for greater efficiency in public bodies and the growing use of the Internet in modern society, have given rise to a new concept called open data. The focal point of this study is open government data (OGD) for the [13], the publication and dissemination of information in the public sector (governments) in the Web, which can be shared in a rational and comprehensible format in a way that can allow it to be recycled in digital applications or in other words, made legible for machines.

Open government and open data are grounded on three pillars: transparency, participation and collaboration. Transparency imposes the responsibility to inform citizens about what exactly the government is doing and what activities it intends to carry out. Participation allows citizen to offer their ideas and skills and thus assists the government to devise more effective and far-reaching policies and also provide more information to society. Collaboration lays stress on the effectiveness of government and encourages closer cooperation between society and the different levels of government.

Until the time when all the initiatives and/or opportunities that the government had for making information available to society (more recently through the use of Internet technology), these questions were traditionally addressed with a view to ensuring transparency, greater control and a better exercise of citizenship. However, with the advent of open data, new opportunities have arisen for society to go beyond these traditional requirements in areas such as the following: developing new goods and services to bring about financial opportunities (through these goods and services), making improvements in social welfare and assisting the government (by providing feedback) through an interaction made possible by these new goods and services.

Authors such as [6, 11] stress that it cannot be expected that simply by opening up new data, governments will begin to produce goods and services and create economic opportunities. For this to take place an attempt must be made to study, understand and suggest improvements or corrections to the proceedings of governments and society in general, with a view to fostering new activities, facing challenges and reducing (or ideally overcoming) unforeseen problems.

Researchers into the environment of information have used the metaphor of the eco-system to stand for the wide and intricate web of relationships between the suppliers, users, data, material infrastructure and institutions. In the opinion of [5], an eco-system is “a system of people, activities, values and technologies in a particular

environment". [5] think that the owners of an eco-system are located in three entities: (a) the policies and practices of governments; (b) the web users, companies and citizens of society; (c) the innovators. These can interact in several ways and influence the way the eco-system evolves: the policies and practices of government can interact with the web users, civil society and companies.

An ecosystem cannot be established in a satisfactory way without the presence of intermediaries. These might be government policymakers or originate in the society where there is a capacity and desire to act in the eco-system of open government data by having access to data and being able to handle and disseminate it. In other words, these data can be made useful for third parties for society through its most varied kinds of representation or organization.

Studies such as those of [8, 9], have underlined the importance of the intermediaries in this ecosystem. Setting out from these studies, these authors recognize that it is important for a wide range of players to take part because several technical activities are needed to handle the available information and make it accessible to the public. They state that a good deal is required to handle the available information and that this entails a close collaboration between the participants of this movement.

For the purposes of this study the authors define the term "intermediaries" as referring to "all the players (in an individual way or representatives of governments and social organizations), who are involved with public data that are released in an open format. They may or may not make use of technological, legal or structural artifacts in their activities. In making use of open data, the intermediaries aggregate value to the data to ensure that they can be understood more easily (and hence have a greater value) for third parties after their intervention". The intermediaries can and must supply goods and/or services with the public database for the government and society/third parties.

2.2 Possible Benefits and Advantages

It can be argued that there are several advantages for society and government in using government data in an open format. As well as having the chance to exercise a greater degree of citizenship, people are given new economic opportunities by having access to public information. The study by [10] estimates that when it is spread to cover areas such as education, transport, health, finance, fuel, and electricity, open data has the potential to add three trillion dollars to the global economy every year.

Governments also benefit from the use of open data. [4, 6], cite the following as possible advantages to governments: greater internal efficiency - by having greater/better access to information, governments can make use of these data to undertake their activities and thus increase their efficiency by using information that is difficult to obtain (even by the governments themselves or other governments); and greater efficiency owing to the feedback provided by society that uses open data can and must "act as an informer" of the government about the possible approach that should be adopted for a better/greater use of public resources. [7] classify these possible benefits and/or advantages in three groups which are as follows: political and social; economic and technical; and operational. These benefits and/or advantages were used for an analysis of the main questions.

2.3 Possible Obstacles and Inhibitory Factors

Some problems can arise in OGDs such as obstacles or possible inhibitors. Governments make great use of ICT; currently, they deploy different technological platforms in which the databases are stored. These databases should be the source of the data from open data projects. Governments have to carry out technical tasks that can be distinguished from the implementation of projects and involve the "conversion" of the data from these bases to formats that are suitable for publication in accordance with the rules of open data.

Another area that serves as a possible obstacle or inhibitory factor concerns the "understanding" of data from society (which is addressed in the issue – the quality and usability of available content). How can one ensure that the data are reliable and that the content is up-to-date? Since the open data must be published in different formats from those stored in the database, the task of "conversion" is a necessary investment when forming the teams [5].

The structure of government bodies can also have an adverse effect on these initiatives since even when there are legal guarantees for the publication of the data, the governments need the operational conditions to carry out this task. It is recognized that there is a need for training and to adapt the culture of the public servers to this end as well as a suitable functional and technological framework for a correct understanding of the system and to meet the legal, technological and structural requirements. [7, 13] classify the obstacles and possible inhibitory factors into six large groups which are as follows institutional, complexity of tasks, use and participation, legislation, quality of information and techniques. These obstacles and inhibitory factors were drawn on for the data analysis.

3 Methodology

In the light of its objectives, this can be characterized as an exploratory study, as defined by [2], who believes that exploratory studies are suited to broadening the researcher's knowledge of a relatively unknown phenomenon and thus providing a greater understanding of it. Another feature of this kind of work is that the researcher does not expect to obtain definitive answers to the problem addressed.

In forming the sample, the authors of the study made use of their findings from a previous work on the "ecosystem" of open government data. Nineteen interviews were conducted, some in person and others via Skype; the interviews were recorded and subsequently transcribed. Four interviews were conducted in Argentina, six in Brazil, five in Mexico and four in Uruguay; all the interviews were carried out between September and December 2014.

The study was based on analysis assumptions with regard to content and a previous categorization of the aims of the research questions according to [1]. This author thinks that conducting an analysis of content is an investigative technique which is employed with a view to obtaining an objective and systematic description of the issue that is being communicated. One of the ways of putting into effect an analysis of content is to categorize the texts that emerge from the data collection. This categorization can be carried out in an a priori way and is grounded on a theoretical basis.

The categories must be valid and there must be a consistent classification of each element. A valid kind of categorization must be significant with regard to the content of the materials which are being analyzed and conform to the goals set out for the study. A qualitative approach was adopted for handling the data analysis which sought to determine what citations or references to previously established categories occurred according to [2]. In Table 1, there is a chart showing the categories that were drawn on for the analysis

Table 1. Categories for analisys.

Category	Characteristics of each category
Profile of the respondent	Terms referring to the profile of the respondent
Action taken in the course of the study	Terms used in making citations to the activities that were carried out
	What activities were carried out? However they carried out? Why were they carried out? What were the reasons for the course of action taken?
Benefits and/or advantages	Particular issues: political, social, economic and operational or technical
Obstacles and/or inhibitory factor	Particular issues: institutional, complexity of tasks, use and participation, legislation, quality of information and techniques
Results attained	Terms that can define the main results that were attained

In Table 2 there is a description of the instrument that was used, together with the purpose of each question and its relation to the categories defined for the analysis. The first part of the interview consisted of questions to determine the profile of the interviewee. The other questions were aimed at finding out about the activities that were carried out, the benefits and/or advantages, obstacles or inhibitory factors regarding the OGD ecosystem, especially with regard to budgetary data.

Table 2. Questions of instrument.

Questions	Purpose of the questions	Category analysis
Identification of the respondent	To find out the name and organization of the respondent	Profile of respondent
How would you define your use of open data and how much time have you spent on open data?	To form a profile of the respondents through their performance	Performance
What are the factors that have led you to play your role (i.e. from the action you have taken) in using open data?	To find out what factors have driven the respondents to take action in this area	Benefits and/or advantages

(*Continued*)

Table 2. (*Continued*)

Questions	Purpose of the questions	Category analysis
What are the main expected benefits (on the basis of your activities) of using open data (for governments and society in general)?	To obtain a better understanding of the driving-force behind the intermediaries in the eco-system	Benefits and/or advantages
What are the main difficulties encountered when using open data?	To find outwhat factors serve as obstacles or inhibitory factors in the eco-system and may, in a general way, impair the performance of the intermediaries	Obstacles and/or inhibitory factors
Is the information available suitable and adequate for carrying out your activities?	To find out the obstacles or inhibitory factors with regard to making information available	Obstacles and/or inhibitory factors
How would you define access to tools and/or resources for taking action to ensure budgetary transparency?	To determine the obstacles or inhibitory factors with regard to resources or tools	Obstacles and/or inhibitory factors
What results stand out from your performance?	This question seeks to find out what results/experiences of the intermediary need highlighting	Results attained a

4 Results and Analysis

In this section there is an examination of the results and respective analysis which follows the order of the questions in the instrument. In this way, the profiles of the respondents are shown together with the way they behaved during the project. Following this, the benefits and advantages are outlined together with obstacles and inhibitory factors.

4.1 The Profile of the Respondents and their Way of Behaving

The respondents basically had professional backgrounds and came from the following areas: public policymaking, economics, law, journalism and information technology Attention should be drawn to the presence of journalist in this group. With regard to the question of their performance, there were actors (intermediaries) of governments, who stressed that their main aim was to support the task of publishing data. On the other hand, there were key players who sought to publish the academic subject of the research they were carrying out.

The intermediaries in the area of information technology stated that they acted in this area to develop tools that could make it easier to access and handle the data. Several actors (intermediaries) stated that they were involved in the observation of public policies and in the context of the interviews, the issues they focused on were

financial matters or the public budget. Another issue that was often cited to justify the work of the intermediaries, were questions related to transparency.

4.2 Interviews: Benefits and Advantages

Table 3 shows the benefits and/or advantages that were revealed by the interviews. In all the tables there is a number beside the factor that is cited, which corresponds to the number of citations given for it. The number of citations of each category is the sum total of the citations of its respective advantages and/or benefits. The results are arranged to include all four countries and each interviewee is able to cite more than one benefit or advantage.

This way of displaying the results can be explained by the fact that the main purpose of the study is to find out the advantages, benefits, obstacles and inhibitory factors in the view of the OGD interviewees who are involved in the ecosystem of budgetary data.

Table 3. Benefits and/or advantages

Category	Benefits and/or advantages cited – number of citations
Political and social (53)	Greater transparency(11); increased participation and self-accountability of citizens (web users) (11); creation of trust in the government (5); public participation in producing data (6); equality of access to the data (1); new government services for the citizens (1); improvement in the formulation of policies (7); creation of new kinds of knowledge in the public sector (5) and new social services (innovations) (6)
Financial (11)	Assists in improving procedures, goods and/or services (1); development of new products or services (3) and the use of collective knowledge (7)
Operational and technical (14)	Improvement in policymaking (10); improvement in decision-making which allows comparisons to be made (1); makes it easier to access and find data (2) and the creation of new databases by combining data (1)

It is clear from the results that the greatest benefits and/or results are linked to political and/or social factors where there are good prospects of bringing about greater transparency. This is one of the three pillars of open government and its citation was either made in a direct form or by mentioning other factors such as: the prospect of encouraging greater participation among the citizens or providing the public with greater access to the data.

Financial gains were also mentioned in this category and there were clear signs of another pillar of open data being found in this category- collaboration. This was clear when the interviewees stated that they were able to derive advantages or benefits from the use of collective knowledge through the use of an intermediary.

The factor highlighted above converge when obtaining operational and technical benefits or advantages in so far as there is a recognition that policymaking can be

improved and that to achieve this a greater participation and collaboration between society and the government is needed through the main players (intermediaries) of these sectors.

With regard to political and/or social issues, it should be noted that some benefits or advantages were cited that could be obtained by governments such as: the creation of trust in the government and the discovery of new knowledge in the public sector. In addition, society can profit from open data in the following areas: new government services and social services for the general public. This confirms the underlying assumption of certain authors that everybody can derive benefits or advantages from the use of public data in an open format.

4.3 Possible Obstacles and/or Inhibitory Factors

Table 4 shows the obstacles and/or inhibitory factors that emerged from the interviews. In all the tables, there is a number beside the factor being cited which corresponds to the number of its citations. The number of citations of each category is the total sum of the mentions of the respective obstacles and/or inhibitory factors. The results have been arranged to include all four countries and each interviewee can cite more than one benefit or advantage.

Several authors stated that they had a lot of difficulties in publishing public data in an open format which originated from the inner structures of the government. This is

Table 4. Obstacles and/or inhibitory factors

Category	Obstacles and/or inhibitory factors – number of citations
Institutional (25)	Lack of uniformity in the policy for publishing data (15); priority given to the interests of organizations to the detriment of the interests of private citizens (5); failure to define the procedures adopted for working and/or interacting with the web users (3) the existence of doubtful standards for working with the web user (2)
Complexity of the task (38)	Duplication of data, the data made available in different ways either before or after the processing, resulting in uncertainty about the source (9); even when the data can be found, the users may not be aware of their potential uses (6); the data formats and datasets are much too complex to be handled and used easily (17); a lack of support for the tools or helpdesk (5) and the focus is on making use of individual datasets whereas the real value may come from a combination of several datasets (1)
Use and participation (40)	A lack of incentives for the user (9); public organizations fail to react to the input of the user (2); the costs are unexpectedly high (1); lack of time to use open data (1); a lack of knowledge about how to use or make sense of the data (15); lack of the necessary capacity to make use of the information (9); no statistical knowledge or understanding of the the potential value or limitations of statistics (3)

(*Continued*)

Table 4. (*Continued*)

Category	Obstacles and/or inhibitory factors – number of citations
Legislation (3)	Written permission required to have access to the data or to reproduce them (2) and the question of the renewal of contracts/agreements (1).
Quality of the information (32)	Lack of information (10); lack of precise information (8); information that is incomplete, only a part of the total picture shown or only conveyed at particular intervals (6); the loss of essential information (1) and the fact that similar data stored in different systems produce different results (7)
Techniques(29)	The data should be in a well defined format which is easily accessible: whereas the data formats are arbitrary, the data format must be strictly defined (1); absence of standards (11); absence of a support to make the data available (1); lack of goals and standards (1); no standard software for the processing of open data (5) and fragmentation of software and applications (10)

confirmed by the interviewees when the following factors had several citations: a lack of uniformity in the policy for publishing data and the fact that priority was given to the interests of organizations to the detriment of those of private citizens.

Other obstacles that were often cited by various authors were linked to the quality of the published information in various ways. This is also confirmed by the interviewees where there were a large number of citations: duplication of data; data being made available in different ways; the data formats and the fact that the datasets are much too complex to handle and use easily; lack of precise information, absence of standards; and the fragmentation of the software and applications. At the same time, attention should be drawn to the low number of citations regarding obstacles or inhibitory factors related to legislation.

There are many references to factors regarding the interests of society and the following had a lot of citations: a lack of incentives for the user; a lack of knowledge about how to use or make sense of the data and a lack of the skills needed to use the information. This is supported by the mention made of the factors related to the ability of the government to work together with society and the difficulties experienced by those wishing to obtain support or clarification of the content made available.

With regard to other factors, the driving-force behind the desire of society to participate in the OGD initiatives, as well as its qualifications for using these data, is to some extent related the factor of participation. These difficulties might be aggravated by a lack of understanding of the content made available and by the skills (some of which are technical) that are needed to have access to the information and be able to handle it.

4.4 Results Attained

Some questions about the instrument seek to find out what results can be attained. The purpose of these questions was to establish the results that were obtained and allow a parallel to be made between the benefits, advantages, obstacles or inhibitory factors and these results and learning experiences.

The main results that were mentioned, corroborate the fact that this issue is still in its early stages. This is because several of the interviewees stated that the principal result was the publication of documents with a view to publishing material about the question which included documents such as books, lists of statistics, and catalogues. Other results are as follows: it has had a small impact but is gradually being formed; there is more discussion of the issue and the community which is being formed around the subject is growing although only gradually; the fact that open data is an important question but still does not have great popular appeal; the fact that the issue is still in its early stages in the countries that are the focal point of this study.

In the previous subject, various advantages and benefits were cited by the interviewees as being susceptible to inclusion in this ecosystem and the results confirm some of these advantages such as: improving policymaking; a rise in the number of accusations of misdemeanors, which might lead to a better use of public resources; a heightened awareness in society of public expenditure and a greater trust in the government.

Some results confirmed that there were obstacles and difficulties, especially with regard to the limited interest and capacity of society to have access to the data or be able to use it. The interviewees cited the following results in support of this fact: a greater concern and awareness of the importance of allowing society to have access to data and the respective use of this data; in some areas there was a greater interest in the use of the data/information made available in this format as, for example, public transport; finally the formation of alliances between organizations of society increases the interest in the subject.

The results that were cited strengthen the need for the formation of networks between governments and society and allows the intermediaries to play a stronger role in this ecosystem. This is because mention was made of the effort and work necessary to obtain some results which must be carried out by the government and key players in society. As a result, certain activities were carried out such as public events and courses aimed at bringing about a greater degree of integration.

5 Final Considerations

Before giving our final thoughts about this work, an attempt will be made to outline some of its limitations and make suggestions for further studies. One drawback of the work which should be mentioned, is the fact that the sample was formed in an intentional or rather, non-probabilistic way which might have caused a degree of bias in the answers. In an attempt to overcome this problem of bias, the authors decided to interview key players (or intermediaries) from different social origins and professional backgrounds, although all of them were linked to the ecosystem of budgeting.

The authors are aware that in future studies it would be valuable to carry out research into other ecosystems of open government data such as in education, and health among other areas, as a means of determining what results are different and/or similar. Forming a sample with intermediaries from other ecosystems will allow greater assistance to be given to public managers in future planning of schemes for the publication of data in an open format.

Undertaking this study has made it possible to confirm some of the advantages and/or benefits that are proclaimed by several authors, especially with regard to the open data which has the potential to exercise transparency and allow society to have more control over the acts of government and thus establish a historic trend. At the same time, it also confirms that few opportunities have been found in the economic field because not many interviewees referred to benefits and/or advantages in this area.

The results make clear that society still has little interest in the issue and the interviewees made several references to this fact. Several corroborative reasons were mentioned such as the difficulties of understanding the context of the data and the need for a higher level of education and culture to include the social players in this ecosystem. It should be underlined that there was also mention of the need for the intermediaries to have knowledge and technical skills (related to the information technology software sector), and a list was made of the technical difficulties of having access to the data or being able to handle it.

It is also evident from the results that the presence and involvement of governments is important since they can derive benefits and advantages from publishing the data in an open format. Governments should also give priority to mitigating factors that might harm the publication of data in a way and thus circumvent obstacles and technical or structural difficulties with regard to the publication of information. Government actions can also help to overcome the limited interest of society in available information, as well as to encourage activities that can lead to an understanding of the material that is published.

It is believed that this study has attained its intended objective of identifying the advantages, benefits, obstacles and inhibitory factors, from the standpoint of the OGDs who are involved in the ecosystem of budgetary data in Latin America. At the same time, the results have made it possible to find answers to the research question since identifying the factors listed above (as well as the results and lessons learned by the key players in the ecosystem of opexn government data) allow the policymakers to profit from the improved conditions for carrying out projects of open government data.

One of the contributions that can be made by the managers is as follows: it is extremely important to create the conditions that can enable organizations like those studies in this work, to be able to carry out their activities and in particular, to have an environment (or an ecosystem) that allows organizations to be combined so that partnerships can be formed in a way that is complementary to their individual capacities. These partnerships can also be formed to obtain benefits or advantages and as a way of overcoming possible obstacles and inhibitory factors through activities that encourage innovation and collaboration, which are the two basic principles of open data.

Acknowledgements. The authors acknowledge the support of the Iniciativa Latino Americana de Datos Abiertos (ILDA, the Latin American Open Data Initiative) and Avina Foundation that made possible this research study.

References

1. Bardin, L.: Content Analysis. 9th edn. 70 editor, Lisboa (2009)
2. Creswell, J.W.: Research project. Qualitative, quantitative and mixed method, 3rd edn. Artmed editor, Porto Alegre (2009)

3. Eaves, D.: The three laws of open government data (2009). http://eaves.ca/2009/09/30/three-law-of-open-government-data/
4. Espinoza, J.F., Recinos, I.P., Morales, M.P.: Datos Abiertos: oportunidades y desafios paraCentroamérica con base en unacadena de valor. Open Data For Development in Latin America and the Caribbean (OD4D), Montevideo, Uruguay (2013)
5. Harrison, T.M., Pardo, T.A., Cook, M.: Creating open government ecosystems: a research and development agenda. Future Internet **4**, 900–928 (2012). http://www.mdpi.com/journal/futureinternet
6. Helbig, N., Creswell, A.M., Burke, B.G., Pardo, T.A., Luna-Reyes, L.F.: Modeling the informational relationships between government and Society. In: Open Government Consultative Workshop – CTG, Albany, NY, US (2012)
7. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. **29**, 258–268 (2012)
8. Kuk, G., Davies, T.: The roles in assembling open data complementarities. In: Thirty Second International Conference on Information Systems, Shanghai, vol. 32 (2011). http://soton.academia.edu/TimDavies/Papers/1216268/The_Roles_of_Agency_and_Artifacts_in_Assembling_Open_Data_Complementarities
9. Mayer-Schoenberger, V., Zappia, Z.: Participation and power: intermediaries of open data. In: 1st Berlin Symposium on Internet and Society, Proceedings of 1st Berlin Symposium. Berlin: Alexander von Humboldt Institutfuer Internet und Gesellschaft (2011). http://berlinsymposium.org/sites/berlinsymposium.org/files/participation_and_power.pdf
10. McKinsey, C.: GI. Open data: Unlocking innovation and performance with liquid information (2013). http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation
11. Prince, A., Jolías, L., Brys, C.: Análisis de la cadena de valor del ecosistema de DatosAbiertos de la Ciudad de Buenos Aires, Montevideo, Uruguay (2013). http://www.princeconsulting.biz/pdf/7.pdf
12. W3C. Improving access to government with the best web use (2009). http://www.w3c.br/divulgacao/pdf/gov-web.pdf
13. Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.S.: Socio-technical Impediments of open data. Electron. J. e Gov. **10**(2), 156–172 (2012). Academic Publishing International Ltd. Zuiderwijk. http://www.ejeg.com

Transparency Dimensions of Big and Open Linked Data

Transparency as Being Synonymous with Accountability and Openness

Ricardo Matheus(✉) and Marijn Janssen(✉)

Delft University of Technology, Delft, The Netherlands
{R. Matheus, M. F. W. H. A. Janssen}@tudelft.nl

Abstract. Although one of the main reasons to open data by governments is to create a transparent government, many initiatives fail to deliver transparency. Making data available does not automatically yield up transparency. Furthermore transparency is an ill-defined concept and understood in different ways. Added on this puzzle, Big Data Analytics are becoming reality in governments and on society due the quantity of open data available and the evolution of techniques and instruments used to analyze data. This paper develops a Big and Open Linked Data (BOLD) Framework identifying categories, dimensions and sub-dimensions that influence transparency. Our framework conceptualizes transparency as a process of data disclosure and usage. Transparency is based on the two major synonymous concept used on literature, namely accountability and openness. Accountability means revealing important details for transparency to control governments financially and operationally, whereas openness reveals details of what, how and why politics took the decision, without revealing important parts of the political game inside government such in military and nuclear area.

Keywords: Big and Open Linked Data (BOLD) · Transparency · Evaluation · Accountability · Openness

1 Introduction

One of the main purposes of opening data by governments is to create transparency [1, 2]. Transparency is often viewed as a condition *sine qua non* for Democracy [3–5]. Since the mid-nineties politicians and governments started to create portals to publish government information including budgeting and statistical data [4]. Some governments created Application Interface Programming (APIs) to enable streaming data. Recently, world watched a boom of Big Data Analytics usage of techniques for some reasons [6] e.g. the sufficient possible quantity of data to be collected by millions of users' devices (smartphones, social networks on Internet), processed at enterprises or opened by government [7], further the fast evolution of hardware to collect, storage, treat and analyze all the data in a small period of time with high velocity of processing [8]. For this we name on this paper Big and Open Linked Data (BOLD).

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 236–246, 2015.
DOI: 10.1007/978-3-319-25013-7_19

All these effort are aimed to create transparency, although the actual contribution to creation of transparency can be challenged [9, 10]. The merely opening of data without providing any user-interface might result in an inability to use, whereas the inclusion of a predefined user-interface and visualization might only result in a biased picture. The more data is opened the higher the information overload. Disclosing open data does not result in transparency, only the actual use of the open data can result in transparency.

A lack of transparency is usually caused by information asymmetry between the government (agent) and the citizens (principal) [11]. Information asymmetry refers to the situation in which one party has more information than the other [12]. In our situation this occurs as government has more information than its constituents. Information asymmetry prevents that all the stakeholders have the same insight on the governmental issues, being able to control the public policies and participate on the public management. The opening of data should overcome the information asymmetry to some extent and provide better insight into the inner part of the government.

Although transparency is an intuitive appealing concept, there is no uniform view on what constitutes transparency. Some authors even define transparency as magical concept [13, 14], whereas others use the term as synonymous with accountability [15–17] and/or openness [9, 18]. In this research we contribute to this discourse by investigating the aspects of transparency and dimensions for opening data to enable transparency. This paper is structured as follows. First we present the research approach that uses literature review and interviews as the primary research instrumented. This is followed by the literature background, which results in the conceptualizations of transparency. Next a Big and Open Linked Data Framework is presented to identify dimensions and subdimensions that influences transparency. This results in two dimensions and theirs sub-dimensions: data dosclosure and data usage. Finally conclusions are drawn.

2 Research Approach

Our aim is to get insights into the factors influencing and conditions for creating transparency. First a literature review of transparency concept was included by searching on the terms “transparency + government” “Transparency + accountability + government” and “Transparency + openness + government”. Papers published in the e-government top journals were reviewed including Government Information Quarterly (GIQ), International Journal of E-Government Research (IJEGR), Transformational Government, People, Processes and Policy (TGPPP), E-Government, an International Journal (EGIJ), International Journal of Public Administration in the Digital Age (IJPADA) and Information Polity (IP). The review was limited to the first 50 papers as presented by Google Scholar and Scopus. The main reason for this limitation is that after the fifth page, the number of publications started to repeat itself and the remaining works was not considered as influential due to the limited number of citations. Another limitation was that some books were not accessible online. Those publications were excluded. A total of 200 publications were found of which 85 publications were considered to be relevant for our research after scanning the papers but only 54 papers were selected as source and reference.

Results having only the keywords inside the list of references were excluded. The literature review revealed two mainstreams: transparency as being synonymous with accountability [15–17] and as synonymous with openness [9, 18]. Next the content of these papers was analyzed for factors affecting and conditions necessary for data disclosure and data usage. The literature review results in transparencyfFramework considering dimensions that influences data disclosure and data usage to promote transparency, identified as a majority flow mechanism to achieve transparency.

3 Views on Transparency

Our literature review revealed that he concept of transparency is complex due to its ambiguity and various usage in the literature and practice [3]. Transparency has been used as a magic concept by governments to improve efficiency [13] or as synonymous with accountability [15–17] and synonymous with openness [9, 18]. In practice the concept of transparency is often misunderstand and more talked about than practiced [19].

In our view transparency is aimed at overcoming the information asymmetry between the government and the public. Transparency refers to the ability for the public to understand the various aspects of government. It is about the ability to see the inner working of the government. This means that who and how decisions are made and what evidence is used are transparent. For this purpose data about the functioning of the government should be released. From this puzzle, two main categories on transparency were identified at literature view: Transparency as synonymous with accountability and transparency as synonymous with openness.

3.1 Transparency as Synonymous with Accountability

Accountability normally involves a relationship between two or more parties, where one party holds the responsibility of performance given certain objectives pre-stablished or planned, taking in consideration public principles such as effectiveness and efficiency use of resources to realize the purposed objective. Accountability implies answerability for one's actions or inactions and the responsibility for their consequences [20]. Accountability means also taken responsibility for decisions. Elections are a case of accountability in governments, where people can judge the past actions of politics after managing the state during some period of time [21]. Accountability concerns the comparison of objectives with the realized performance and deviations [16].

Some actions, which are often considered as part of transparency, are in fact actions necessary for accountability and keeping politicians and public officials accountable. Many open data are just a publicity of data or the published data cannot be used for accountability, as its characteristics are not suitable for this. For example too low quality or only providing insight into one aspect. Transparency as accountability is also used to identify when it is possible to enable anti-corruption in government [22, 23]. For the practice perspective, the Transparency International [24] has lead its objective

of transparency toward the anti-corruption goal, identifying who has not a suitable work within public management. In their index the term transparency is used to advocate its objective, the anti-corruption. Nevertheless, in our view transparency is the way to enable anti-corruption tools. In conclusion, information through transparency of governments is the raw material to enable accountability or anti-corruption tools. Accountability uses data/information [25] and do something with the information that publicity created [16]. Yet accountability does not need complete transparency. Some activities might be hidden, as they are no needed for being accountable. Only surrogates are published which are necessary to keep one accountable. No knowledge of the inner working of the public system needs to be published. Those parts remain hidden and are not transparent. From this we conclude that although accountability and transparency are overlapping they are also distinct.

3.2 Transparency as Synonymous with Openness

After being sworn in the United States of America, the President Barack Obama created the Memorandum of Open Government with the aim of "creating an unprecedented level of openness in Government" [26]. One of the underlying goals was to create an open and transparent government. Openness does not represent automatically the result of increasing transparency in governments, however, it influenced the creation of open government data portals and legislations as a freedom of information act. For example a box can be open, but still it might not be transparent and you cannot see inside it. On the other hand openness might be necessary condition for transparency. If the system is closed there cannot be any transparency. From this we conclude that although openness and transparency are overlapping but distinct concepts.

Scientific literature points out that openness is also close to open government initiatives [27, 28]. From this, scientific literature and practical people are facing an operational and theoretical definition toward openness as Open Government concept. For governments it is not an initial stage of the word and the concept, however, for military/civil usage of nuclear weapon/energy [29, 30] and the finances [28], openness is not also on initial phase. Both of them can be inspired on how processes can be at same time transparent, open and not show evidence of core business or private information, what basically is citizens' need to do accountability and politics for governance without revealing important parts of the political game inside government.

4 Towards a Transparency Framework

4.1 Basic Framework

Basically the BOLD has three-steps: Collecting data, from data internal databases, spreadsheets, document files, sensors spread over a city or a social network on Internet. Secondly, storage of data, that requires advanced and unique data storage, management [31] and thirdly analysis and visualization technologies [6]. In the opening of data there are two important stakeholders the publishers and users of open data. The publishers and users often are unaware of each other needs and encounter different challenges and

barriers [32]. The data publishers' main activity is the disclosure of data, which is necessary before data can be used. Yet only releasing data does not result in any transparency. Only the actual use of data results in transparency. Both steps are influenced by a large amount of factors. In conclusions, transparency can only be created when both data disclosure and usage happens. In Fig. 1 the basic transparency framework in the form of these two essential activities are presented. Hereafter we delve into the details of factors and conditions impacting data disclosure and usage.

Fig. 1. - Basic transparency framework

4.2 BOLD Transparency Framework

From the basic framework, was possible to identify that each dimension could be deep described with sub-dimensions. The Fig. 2 reveals it and each dimensions identified was deeply described on the Sect. 4.3.

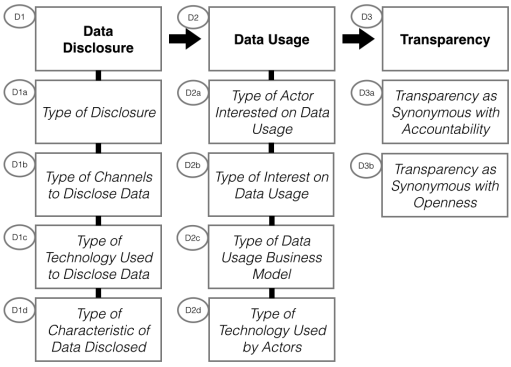


Fig. 2. The BOLD Transparency Framework

4.3 Data Disclosure Category

Being transparent requires the disclosure of data. The disclosure of data is a condition and a first principle for creating transparency. Yet simple making data online is not sufficient. Meta-data about the information quality, the way information is disclosed influences the actual transparency. Four dimensions were identified as follow and summarized at Fig. 2.

A. Type of Data Disclosure. The literature review [33] identified that disclosure of data and information in government occurs when there is Proactive dissemination by the government, Release of requested materials by the government, Public meetings

and Leaks from whistleblowers. The disclosure of data prompts two "types of data disclosure", a first sub-dimension of dimension "data": (i) formal and (ii) informal. Furthermore it was identified the existence of different channels to disclose, sometimes, the same kind of data, or eventually, on different formats and conditions.

B. Type of Channels to Data Disclosure. On the dimension "types of channels to disclose data" the following types of channels were identified: (i) transparency portals [34], (ii) freedom of information access (FOIA) using all kind of channels [35], (iii) open government data portals [36], (iv) governmental portals, (v) dashboard of services advertising, (vi) outdoors with accountancy expenditures data in public works, (vii) public financial statements (newspaper, (viii) paper based at blackboards on City Hall, (ix) Internet based on portals), (x) call center to provide information and (xi) call center to provide access and demand to public services [37]. The channels aforementioned can use different kinds of technologies and some of them were explicated such as public financial statements, with newspaper, paper based at blackboards on City Hall and Internet based via portals.

C. Type of Technology Used to Disclose Data. The third dimension, "Type of technology used to disclose data", is part of the dimension "data" disclosure and a list of technology identified on literature is presented by: (i) politics discourses and civil servant responses, (ii) printed based (paper, newspapers, outdoors of public works and services, signs, etc.), (iii) electronic formats of data (static web portal, downloadable files, etc.) [38] and (iv) real-time electronic accessable data (direct databases access and APIs) [39].

D. Type of Characteristic of Data Disclosed. This implies on the fourth dimension "type of characteristic of disclosed data", that comprehends factors and conditions of data, such as the "quality of data", taking in consideration the type of disclosure, type of channels and technology used and characteristics identified at literature. It is important to highlight the types identified were not deeply discussed at this point and will be approached on next publication, deepening the dimensions and sub-dimensions found here and presented below: (i) data accuracy [40], (ii) data timely [41], (iii) data acessibility [42], (iv) data completeness [43], (v) data security [44], (vi) data trustiness, (vii) data free [26], (viii) data documentation, (ix) data permanently and history, (x) data primarily [41], (xi) data metadata and interlinked [45], (xii) data non-proprietary and non-discrimnatory [41, 46], (xiv) data license-free [26, 41], (xv) data machine processability [41, 46], xvi) portal simple language [47], (xvii) open data policy and license [26, 41].

4.4 Data Usage Category

Only publishing data does not create transparency. The second step of the framework flow of transparency is the Data Usage, in which the public usage data to address solutions to solve theirs interests through the best technology they consider. Taking this principle, the dimension "Data Usage" has four sub-dimensions, factors and conditions as follow and summarized at Fig. 2.

A. Type of Actor Interested on Data Usage. Anyone or any organization need to use data for some specific interest, running a determined business model and with a chosen technology to collect, treat and analyze the data. The scientific and practical literature [48, 49] identified five actors that have been using transparent data, whatever the types already described at Data Usage category and its dimensions: (i) Academics, (ii) Enterprises, (iii) Governments, (iv) Journalists and (v) Organized Civil Society.

B. Type of Interest on Data Usage. From the actors that use disclosed data, it is necessary to comprehend their interest to use the data. The scientific and practical literature [37, 49, 50] identified four types of interest on data usage dimension: (i) Service Delivery, (ii) Accountability, (iii) Advocacy and (iv) Participation.

C. Type of Data Usage Business Model. To sustain the data usage is necessary a business model. The scientific and practical literature [37, 49, 50] identified six types of data usage business model dimensions: (i) Big Data Analysis in governments, (ii) Governmental Portals and procedures for participation, (iii) Governmental portals for social control (accountability), (iv) Organized Civil Society portals and procedures for participation, (v) Organized Civil Society Data Visualization portals to intermediate relationship between governments and civil society and (vi) Private applications to improve service delivery ran by advertising or supported by civil society organizations.

D. Type of Technology Used by Actors. Taking in consideration that actor has an interest and a business model to use the data, the last type of dimensions is technology. The scientific and practical literature [7, 51] identified six types of technology that can be used: (i) Computer programming languages, (ii) Data Visualization, (iii) Geography Coding and Mapping, (iv) Networking Analysis, (v) Business Intelligence and (vi) Data Mining.

4.5 A Framework with the Dimension of Transparency

From the BOLD Transparency Framework, is possible to addres the Table 1 - BOLD Framework Summary Dimensions and Sub-Dimensions.

Table 1. - BOLD framework summary dimensions, sub-dimensions and types of specifications and characteristics

# ID	Name of category and dimensions	Sub-dimensions identified
D1	Data disclosure	–
D1a	*Type of disclosure*	*Formal* (Proactive dissemination by the government, Release of requested materials by the government, Public meetings) and *Informal* (Leaks from whistleblowers)
D1b	*Type of channels to disclosure data*	Transparency portals freedom of information access (FOIA) using all kind of channels open government

(*Continued*)

Table 1. (*Continued*)

# ID	Name of category and dimensions	Sub-dimensions identified
		data portals, governmental portals, dashboard of services advertising, outdoors with accountancy expenditures data in public works, public financial statements (newspaper, paper based at blackboards on City Hall, Internet based on portals), call center to provide information and call center to provide access and demand to public services
D1c	*Type of technology used to disclose data*	politics discourses and civil servant responses, printed based (paper, newspapers, outdoors of public works and services, signs, etc.), electronic formats of data (static web portal, downloadable files, etc.) and real-time electronic accessible data (direct databases access and APIs)
D1d	*Type of characteristics of data disclosed*	data accuracy, data timely, data acessibility, data completeness, data security, data trustiness, data free, data documentation, data permanently and history, data primarily, data metadata and interlinked, data non-proprietary and non-discriminatory, data license-free, data machine processability, portal simple language, open data policy and license
D2	Data usage	–
D2a	*Type of actor interested on data usage*	Academics, Enterprises, Governments, Journalists and Organized Civil Society
D2b	*Type of interest on data usage*	Service Delivery, Accountability, Advocacy and Participation
D2c	*Type of data usage business model*	Big Data Analysis in governments, Governmental Portals and procedures for participation, Governmental portals for social control (accountability), Organized Civil Society portals and procedures for participation, Organized Civil Society Data Visualization portals to intermediate relationship between governments and civil society and Private applications to improve service delivery ran by advertising or supported by civil society organizations
D2d	*Type of technology used by actors*	Computer programming languages, Data Visualization, Geography Coding and Mapping, Networking Analysis, Business Intelligence and Data Mining

5 Conclusions

Transparency is a multi-facetted concept and stakeholders give different meanings to the concept. An important contribution is the identification of two transparency concepts that are often used in the literature. One concept is accountability and the other

openness. Both are overlapping with transparency, but are distinct concepts. We define transparency as the level of insight into functioning of the government. For this purpose data should be disclosed.

Although open data disclosure is a condition for transparency only the actual use can result in transparency. Hence a framework consisting of open data disclosure and use resulting in transparency was proposed where dimensions and sub-dimensions were revealed.

The identified dimensions help to understand what influences the synonymous types of transparency concept. We recommend combining information quality literature with the framework presented in this paper. Furthermore we suggest to filter the dimensions identified in this paper and determine its magnitude of influence on transparency. The next research paper will provide a case study based on the BOLD Transparency Framework identified to find refine the dimensions and sub-dimensions.

References

1. Bertot, J.C., Jaeger, P.T., Grimes, J.M.: Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Gov. Inf. Q. **27**, 264–271 (2010)
2. Dawes, S.S., Helbig, N.: Information strategies for open government: challenges and prospects for deriving public value from government transparency. In: Wimmer, M.A., Chappelet, J.-L., Janssen, M., Scholl, H.J. (eds.) EGOV 2010. LNCS, vol. 6228, pp. 50–60. Springer, Heidelberg (2010)
3. Bersch, K., Michener, G.: Conceptualizing the quality of transparency. In: 1st Global Conference on Transparency Rutgers University, Newark (2011)
4. CETIC: Survey on the use of information and communication technologies in Brazil : ICT households and enterprises 2013. Brazilian Internet Steering Committee - Regional Center for Studies on the Development of the Information Society (Cetic.br), Sao Paulo (2014)
5. Hamilton, A., Madison, J., Jay, J., Pole, J.R.: The Federalist. Hackett Publishing, Indianapolis (2005)
6. Russom, P.: Big data analytics. TDWI Best Practices Report, Fourth Quarter (2011)
7. Chen, Y.-C., Hsieh, T.-C.: Big data for digital government: opportunities, challenges, and strategies. Int. J. Public Adm. Digital Age (IJPADA) **1**, 1–14 (2014)
8. Salvador, E., Sinnott, R.: A cloud-based exploration of open data: promoting transparency and accountability of the federal government of Australia. In: 1st Symposium on Information Management and Big Data, p. 22 (2014)
9. Bannister, F., Connolly, R.: The trouble with transparency: a critical review of openness in e-Government. Policy Internet **3**, 1–30 (2011)
10. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manage. **29**, 258–268 (2012)
11. Eisenhardt, K.M.: Agency theory: an assessment and review. Acad. Manage. Rev. **14**, 57–74 (1989)
12. Jensen, M.C., Meckling, W.H.: Theory of the firm: managerial behavior, agency costs, and ownership structure. In: Brunner, K. (ed.) Economics Social Institutions, vol. 1, pp. 163–231. Springer, The Netherlands (1979)
13. Ward, S.J.: The magical concept of transparency. In: Ethics for Digital Journalists: Emerging Best Practices, p. 45 (2014)

14. Baume, S., Papadopoulos, Y.: Bentham revisited: transparency as a "magic" concept, its justifications and its skeptics. In: Transatlantic Conference on Transparency Research, University of Utrecht, 7 June 2012 (2012)
15. Armstrong, E.: Integrity, transparency and accountability in public administration: recent trends, regional and international developments and emerging issues. United Nations, Department of Economic and Social Affairs, pp. 1–10 (2005)
16. Naurin, D.: Transparency, publicity, accountability-the missing links. Swiss Polit. Sci. Rev. **12**, 90 (2006)
17. Fox, J.: The uncertain relationship between transparency and accountability. Dev. Pract. **17**, 663–671 (2007)
18. Bugaric, B.: Openness and transparency in public administration: challenges for public law. Wis. Int'l LJ **22**, 483 (2004)
19. Hood, C., Heald, D.: Transparency: The Key To Better Governance? Oxford University Press, Oxford (2006)
20. Roberts, N.C.: Keeping public officials accountable through dialogue: resolving the accountability paradox. Public Adm. Rev. **62**, 658–669 (2002)
21. Przeworski, A., Stokes, S.C., Manin, B.: Democracy, Accountability, and Representation. Cambridge University Press, Cambridge (1999)
22. Ball, C.: What is transparency? Public Integrity **11**, 293–308 (2009)
23. Florini, A., Birdsall, N., Flynn, S., Haufler, V., Lipton, D., Morrow, D., Sharma, S.: Does the invisible hand need a transparent glove? the politics of transparency. In: World Banks Annual Conference on Development Economics, pp. 163–184 (2000)
24. Index, C.P.: Transparency International (2007)
25. Michener, G., Bersch, K.: Conceptualizing the quality of transparency. In: paper apresentado na 1ª Conferência Global sobre Transparência, ocorrida na Rutgers University, Newark, em maio de (2011)
26. Obama, B.: Transparency and open government. Memorandum for the heads of executive departments and agencies (2009)
27. Coglianese, C.: The transparency president? the Obama administration and open government. Governance **22**, 529–544 (2009)
28. Birchall, C.: Introduction to 'secrecy and transparency': the politics of opacity and openness. Theor. Cult. Soc. **28**, 7–25 (2012)
29. Gonçalves, M.E.: Transparency, openness and participation in science policy processes. In: Pereira, A.G., Vaz, S.G., Tognetti, S. (eds.) Interfaces Between Science and Society, pp. 180–188. Greenleaf Publishers, Sheffield (2006)
30. Goncalves, O.D.: Openness and transparency, stakeholder involvement. In: Proceedings of an International Conference on Effective Nuclear Regulatory Systems: Further Enhancing the Global Nuclear Safety and Security Regime (2010)
31. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**, 1165–1188 (2012)
32. Zuiderwijk, A., Janssen, M.: Barriers and development directions for the publication and usage of open data: a socio-technical view. In: Gascó-Hernández, M. (ed.) Open Government, vol. 4, pp. 115–135. Springer, New York (2014)
33. Piotrowski, S.J., Van Ryzin, G.G.: Citizen attitudes toward transparency in local government. Am. Rev. Public Adm. **37**, 306–323 (2007)
34. Matheus, R., Ribeiro, M.M., Vaz, J.C., de Souza, C.A.: Using internet to promote the transparency and fight corruption: Latin American transparency portals. In: Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance, pp. 391–392. ACM (2010)

35. Jaeger, P.T., Bertot, J.C.: Transparency and technological change: ensuring equal and sustained public access to government information. Gov. Inf. Q. **27**, 371–376 (2010)
36. Zuiderwijk, A., Janssen, M.: Open data policies, their implementation and impact: a framework for comparison. Gov. Inf. Q. **31**, 17–29 (2014)
37. Matheus, R., Vaz, J.C., Ribeiro, M.M.: Open government data and the data usage for improvement of public services in the Rio de Janeiro City. In: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, pp. 338–341. ACM (2014)
38. Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.S.: Socio-technical impediments of open data. Electron. J. e-Gov. **10**, 156–172 (2012)
39. Wong, A., Liu, V., Caelli, W., Sahama, T.: An architecture for trustworthy open data services. In: Damsgaard Jensen, C., Marsh, S., Dimitrakos, T., Murayama, Y. (eds.) IFIPTM 2015. IFIP AICT, vol. 454, pp. 149–162. Springer, Heidelberg (2015)
40. Loiacono, E.T., Watson, R.T., Goodhue, D.L.: WebQual: a measure of website quality. Mark. Theor. Appl. **13**, 432–438 (2002)
41. Obama, B.: Making Open and Machine Readable the New Default for Government Information. Executive Order, Washington DC (2013)
42. Perdue, R.R.: Internet site evaluations: the influence of behavioral experience, existing images, and selected website characteristics. J. Travel Tourism Mark. **11**, 21–38 (2001)
43. Dawes, S.S.: Stewardship and usefulness: policy principles for information-based transparency. Gov. Inf. Q. **27**, 377–383 (2010)
44. Pfleeger, C.P., Pfleeger, S.L.: Security in Computing. Prentice Hall Professional Technical Reference, Upper Saddle River (2002)
45. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D.: Tabulator: exploring and analyzing linked data on the semantic web. In: Proceedings of the 3rd International Semantic Web User Interaction Workshop, Athens, Georgia (2006)
46. Strong, D.M., Lee, Y.W., Wang, R.Y.: Data quality in context. Commun. ACM **40**, 103–110 (1997)
47. Mendonça, D., Jefferson, T., Harrald, J.: Collaborative adhocracies and mix-and-match technologies in emergency management. Commun. ACM **50**, 44–49 (2007)
48. Zuiderwijk, A., Janssen, M., Poulis, K., van de Kaa, G.: Open data for competitive advantage: insights from open data use by companies. Proceedings of the 16th Annual International Conference on Digital Government Research, pp. 79–88. ACM, Phoenix (2015)
49. Zuiderwijk-van Eijk, A., Janssen, M.: Participation and data quality in open data use: open data infrastructures evaluated. In: Proceedings of the 15th European Conference on e-Government, pp. 18–19, June 2015, Authors version. ACPI, Portsmouth (2015)
50. Matheus, R., Ribeiro, M.M., Vaz, J.C.: New perspectives for electronic government in Brazil: the adoption of open government data in national and subnational governments of Brazil. In: Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, pp. 22–29. ACM (2012)
51. Gray, J., Chambers, L., Bounegru, L.: The Data Journalism Handbook. O'Reilly Media Inc., Sebastopol (2012)

Open Data Landscape: A Global Perspective and a Focus on China

Charles Shen(✉), Zainab Riaz, Madhuri S. Palle, Qiurui Jin, and Feniosky Peña-Mora

Advanced ConsTruction and InfOrmation techNology (ACTION) Laboratory, Civil Engineering and Engineering Mechanics, Columbia University, New York, USA
charles@cs.columbia.edu

Abstract. Governments are producing significant public data that, if made open, is expected to create enormous social and commercial value as well as improve the civil governance. Unleashing the true power of open public data requires a much better understanding of its ecosystem than is known currently. This paper surveys the global open data landscape by taking into account the Open Data Barometer (ODB) ranking system and its three sub-indexes - readiness, implementation and impact. These indexes are compared and analyzed on the basis of income levels of the ODB ranked countries. Finally, using air quality open data, data availability in developing countries like China is compared with countries of better practices such as UK and US. The comparison helps in understanding the current situation and barriers in opening data in China.

Keywords: Open data · Data availability · Air quality data

1 Introduction

Open data is data that can be used, re-used and distributed amongst the people without any legal, technological or social restriction [1]. It is, therefore, becoming a philosophy where data is accessible to the public for free. This approach induces a sense of accountability and transparency by building a bridge between the people and the government/organizations. Furthermore, it seeks to move beyond transparency, towards a problem solving platform in which open data can become a stepping stone to: drive more effective decision-making and efficient service delivery; spur economic activity; and empower citizens to take an active role in improving their own communities [2].

Researchers have looked at open data from many different perspectives. Some researchers focus on the relevant initiative in individual countries [3, 4] or cities [5]. Other researchers have revealed the political, social and economic impact of open data [6, 7]. Another interesting topic on open data concerns its business aspect. Hartmann et al. [8] looked at the types of business models amongst companies relying on data as their key business resource, and discussed capturing value through data driven business models. Magalhaes and Manley [9] examined 500 US firms that use Open Government

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 247–260, 2015.
DOI: 10.1007/978-3-319-25013-7_20

Data (OGD) and classified them into three categories of business models: enablers, facilitators and integrators. Success stories of many practitioners of open data companies and governments have also been presented [2, 10]. In addition, integrating OGD into the Web of Linked Data has also been investigated [11] where Linked Data describes a method of publishing structured data such that it becomes more useful through semantic queries.

2 Research Motivation and Approach

The motivation behind this particular research is to tackle urban challenges from an open data perspective. According to UN [12], the percentage of population residing in urban areas is expected to increase from 30 % in 1950 to 66 % in 2050. Among the four groups of countries with different income levels [13], countries that are experiencing the fastest pace of urbanization since 1950 are upper-middle-income countries such as Brazil, China and Mexico. Recently, the Chinese government has released a plan for integration of the Beijing-Tianjin-Hebei (Jing-Jin-Ji) regions, which together is the largest mega-region covering 216,000 sq. km and affecting more than 100 million people [14]. The unprecedented urbanization poses ever-increasing sustainable development challenges to cities and the newly urbanized population.

Our research on Open Data for Sustainable Urbanization (ODSU) aims to tackle the urban challenges of these developing countries with a focus on China. In particular, we want to understand whether and how the open data ecosystem can play a role towards a more sustainable and efficient urbanization process. This research is planned in three phases. In the first phase the global open data landscape is surveyed using secondary data. On global scale, countries are categorized based on four different income levels; at the county scale we focused particularly on China, US and UK. In the ongoing second phase of our research, we are collecting and analyzing an extensive list of open data urban applications around world's major cities, identifying the best practices and evaluating their impact on urbanization efficiency. In the third phase of this research, we plan to propose and implement appropriate open data use cases in China's Jing-Jin-Ji area as a pilot study.

This paper, however, only looks into the first phase of the planned research where it attempts to survey the global landscape of open data by taking into account the Open Data Barometer (ODB) ranking index and its associated sub-indexes namely, readiness, implementation and impact [15]. An analysis is performed in order to understand the relationship between these indexes and the income levels of ODB ranked countries. Finally, to realize the current open data situation in China, a comparison is provided with the trendsetters in the ODB ranking index i.e. UK and US.

2.1 Choice of ODB

A number of open data benchmarks have been developed such as World Bank's Open Data Readiness Assessment (ODRA) [16], World Wide Web Foundation's Open Data Barometer (ODB) [15], Open Knowledge Foundation's Open Data Index (ODI) [17]

and Capgemini Consulting's Open Data Economy (ODE) [18], to name few global and widely used benchmarks. However, each of these benchmarks serves a different purpose and focus. Susha et al. [19] suggest that ODB provides a more comprehensive perspective since it not only includes measures at various stages like readiness, implementation, and impact but also highlight the importance of involvement of major stakeholders and challenges throughout the open data process.

According to the authors [19], the ODB offers an insightful analysis of the entire chain (readiness, implementation, and impacts) and is a goal-oriented measure that can be used to realize how to modify implementation so as to accomplish a particular impact (economic, social, or political). However, the authors also suggest that most open data benchmarks (except for ODRA of the World Bank) produce results that are generic and ambiguous and the ranks of countries should not be expected to convey a strictly numeric position of a country but rather an approximation of reality. They specifically consider ODB more argumentative when it comes to open data diffusion.

For this particular research ODB is selected since it offers a snapshot of open data diffusion worldwide. Moreover, the research objective is to see the role of open data when dealing with unprecedented urbanization challenges particularly in developing countries like China. ODB, in addition to readiness, also offers perspective on the implementation and impact stages, which are considered useful when it comes to understanding the urban applications in developed countries and to develop guidelines for countries like China.

Furthermore, UK and US have been identified for comparison with China in this research since China is an example of the upper-middle-income countries with the fastest urbanization rate while US and UK are two high-income countries with high urbanization rate. Also, these two countries have been consistently highlighted as trendsetters by all major open data indexes. UK is ranked highest by the ODB [15], the ODI [17], the PSI Scoreboard [20] and identified as one of the trendsetters by Open Data Economy [18]. The US, in addition to being ranked 2nd by ODB and 8th by ODI, ranks highest in terms of data availability and data portal usability [18].

3 Open Data Overview

The Open Data Barometer (ODB) ranking system is a part of the World Wide Web Foundation's work on the common assessment methods for open data [15]. The weightage of each sub-index is given in Table 1. Using the ODB scores heat maps[1] are developed and presented (Figs. 1, 2 and 3) where the country colour depicts whether it's sub-index is high, moderate or low. The heat map was made so as to divide the total number of countries equally into 3 different layers so as to allow proper comparison. The lightest layer represents those countries in the bottom 1/3rd of the index ranking, the moderate layer represents those in the middle 1/3rd, and the darkest layer represents those countries in the top 1/3rd of the index ranking.

[1] Tool acknowledgement: www.knoema.com.

Table 1. ODB sub-indexes and their weightage (Source: ODB second edition 2014)

ODB Sub-Index	Sub-Index components and weightage
Readiness (25 %) The readiness index measures how the country in consideration is using existing open data and the extent to which open data is available.	Government (33.3 %)
	Entrepreneurs and Business (33.3 %)
	Citizens and Civil Society (33.3 %)
Implementation (50 %) The implementation sub-index is dependent on: Innovation, Social Policy and Accountability	Accountability Dataset Cluster (33.3 %)
	Innovation Dataset Cluster (33.3 %)
	Social Policy Dataset Cluster (33.3 %)
Impact (25 %) The impact sub-index is dependent on three categories: political, social and economic	Political (33.3 %)
	Economic (33.3 %)
	Social (33.3 %)

As seen on the heat map (Fig. 1) North America, a large part of Europe, Australia, Japan and South Korea have strong readiness sub-indexes. This means that these countries have strong government open data initiatives along with entrepreneur, business and citizen participation. By plotting the scored implementation sub-index of all 86 countries on a heat map, it can be observed that a few more countries fall under the high range on the heat map for the implementation index including Russia, Chile and Brazil (Fig. 2). These countries have high implementation index even though they do not have a high readiness sub-index, which leads us to question the dependency between these two sub-indexes. Finally for the impact sub index (Fig. 3), it can be seen that the only countries that seem to have a high impact sub-index are those in Europe, US, Canada and New Zealand

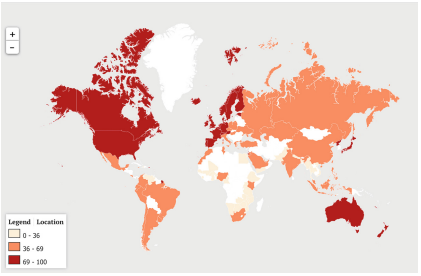


Fig. 1. Readiness sub-index heat map

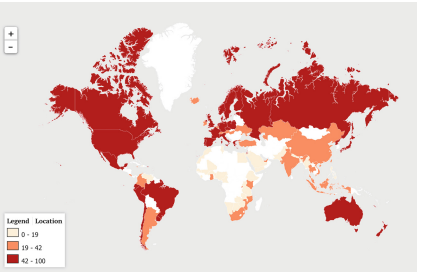


Fig. 2. Implementation sub-index heat map

A general realization here is that many countries have a lower impact sub-index compared to their other two sub-indexes, questioning the impact that open data has on their political, economic and social standings of these countries. Moreover, another observation from the heat maps

is that sub-indexes need not show complete dependency on one another. Following are few examples that reinforce this interpretation:

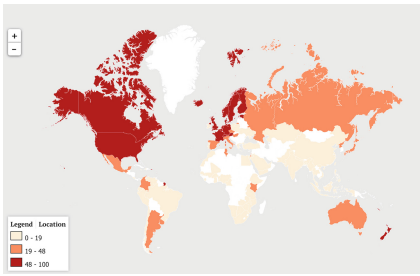


Fig. 3. Impact sub-index heat map

- Brazil has a moderate readiness sub-index, a high implementation sub-index and a low impact sub-index;
- Australia has a great readiness and implementation indexes but not too strong impact index doesn't;
- China and India along with some more other Asian countries have decent readiness and implementation indexes, but low impact indexes;
- Russia, Ecuador and Chile have moderate readiness sub-index but strong implementation index.

4 Open Data Overview Based on Income

The 86 countries listed in the ODB ranking system are divided into four levels of income categories as per World Bank [13] - low, lower-middle, upper-middle and high. Heat Maps for ODB score of all countries in all four categories are developed and presented (Figs. 4, 5, 6 and 7).

The low-income group consists of 16 countries out of the 86 ODB ranked countries. The average ODB score of this group is 11.69. As can be seen from the heat map (Fig. 4) mostly the countries have a low ODB score in this category with a few exceptions in Africa and Indian sub-continent. The lower-middle income group (Fig. 5) consists of 14 countries with an average ODB score of 17.66. As seen from the heat maps (Figs. 4, 5) the low and lower-middle income-ODB rank categories constitute mostly of countries from Asia and Africa.

The Upper-Middle income layer comprises of 21 countries with an average ODB score of 28.57. From the heat map (Fig. 5), it can be seen that the countries of this group that are part of Asia and Africa generally have a lower ODB score than those that fall under South America and North America.

The high-income group consists of 35 countries with an average ODB score of 57.14. An interesting observation is that all countries in the EU are not dark blue on the map. It can be seen that even though EU in general has a high open data standing [15], these practices are not standardized across the region. Furthermore, high ODB scores of countries such as the US and UK result in an overall data skew of this income class.

Figure 8 compares the lowest, average and highest ODB scores for a particular income region. The graph shows that the most significant rise is from the upper middle layer to the high layer - the low, average and high bars increase by around 50 %. Another interesting trend is observed when low to lower-middle and lower-middle to upper-middle layers are compared. Here the average and high values seem to increase by around 30 % which is also a significant figure. Therefore, on a general note, it can be

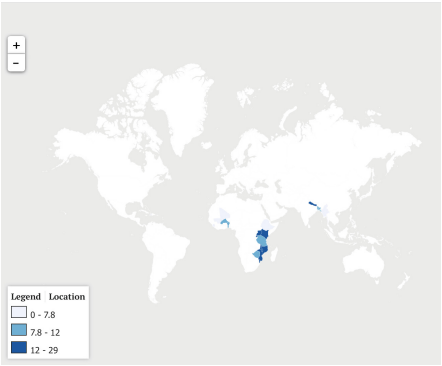


Fig. 4. Heat map for ODB scores of low income countries

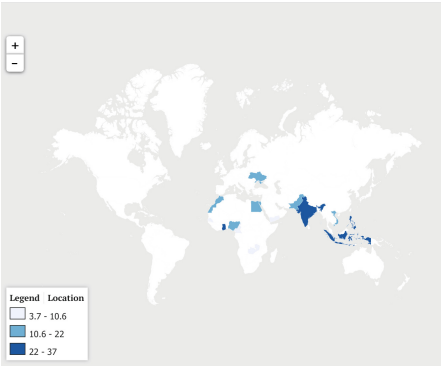


Fig. 5. Heat map for ODB scores lower-middle income countries

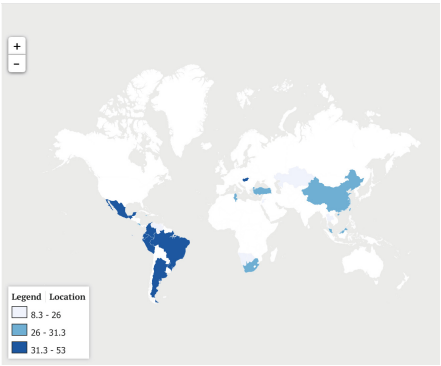


Fig. 6. Heat map for ODB score of upper-middle income countries

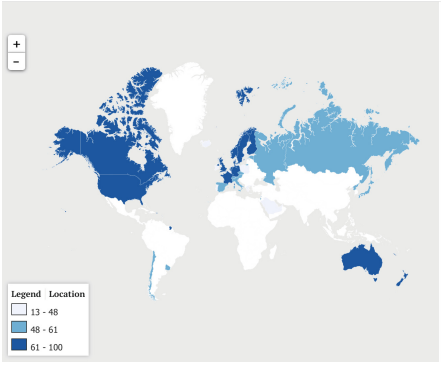


Fig. 7. Heat map for ODB score of high income countries

concluded that as the income increases, it is likely that ODB scores of the countries in that category also increase. This graph consists of the overall ODB score and makes an interesting case to look at the three sub-indexes separately to understand the sharp rise from the upper-middle to high income class.

Next, the three separate sub-indexes are analyzed for individual trends. The first sub-index, the readiness index (Fig. 9), has a uniform increase across all the income classes for the low, average and high values. As can be seen from the graph in Fig. 9, most changes are in the range of 33–47 %. Also, the average rise in all income levels is considerably uniform as compared to the overall ODB index analysis. Therefore, it can be concluded that the sharp rise from the upper-middle to the high class of the overall ODB scores is not dependent on the readiness sub-index.

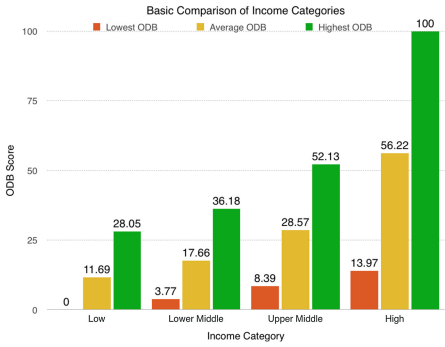


Fig. 8. Basic comparison of income categories

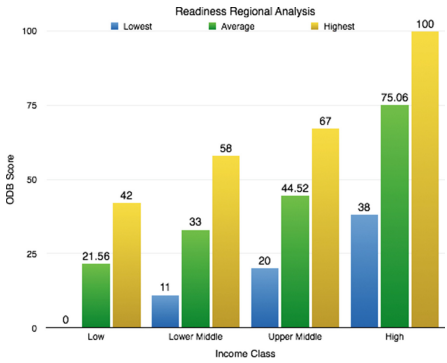


Fig. 9. Readiness regional analysis

Now, observing the graph for the implementation sub-index (Fig. 10) shows that the lowest, average, and highest values across all the income categories are increasing, however, not uniformly. In fact, the rise from the upper-middle class to the high class appears to be very similar to that of the increase seen in the overall ODB analysis earlier. Since the implementation sub-index weighs 50 % of the entire ODB index, it can be concluded that this jump plays an important role in the upsurge portrayed in the overall ODB analysis. Moreover, it can also be observed from Fig. 10 that the rise in the lower category of income is not following the same trend as the overall ODB and the readiness sub-index pattern (both have a constant increase in all levels). In fact, the percentage increase of the implementation index actually decreases for the lower-middle category. Also, from the graph it can be noticed that some countries, despite their reasonably high income, have not implemented open data as efficiently as one would expect them to do so. In this income group, 9 countries have a sub-index of less than 40, which is 25 % of the total number of countries in the class.

The impact sub-index (Fig. 11), on the other hand is very different from the other two sub-indexes. Every income class has at least one country with a zero sub-index. This fact, along with the minimal averages, leads one to believe that this is by far the weakest index. Although countries are showing to have reasonable readiness and implementation sub-indexes, their impact sub-index is below par, proving that more open data initiatives are needed in the economic, social and political sectors. The only similarity with the other two sub-indexes is the surge from the upper-middle income class to the high income class. Although the impact sub-index weighs only 25 % of the total ODB score of a country, this rise in the impact index plays a considerable role in the jump seen in the overall ODB analysis because the percentage increase is extremely large. Therefore, it can be concluded that the implementation sub-index and the impact sub-index are responsible for the increase in the overall ODB scores from the upper-middle income class to the high income class.

In order to understand the overall landscape of the ODB ranked countries along with the representation of the three sub-indexes a line graph (Fig. 12) is developed. Here, the overall ODB country rank, 86 in total, is plotted on the x-axis whereas the

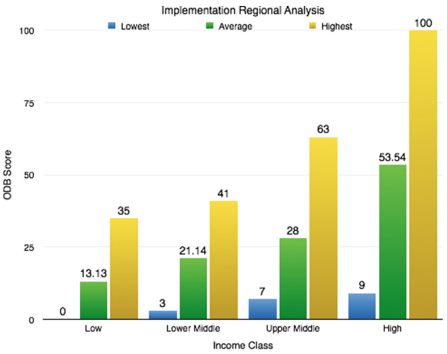


Fig. 10. Implementation regional analysis

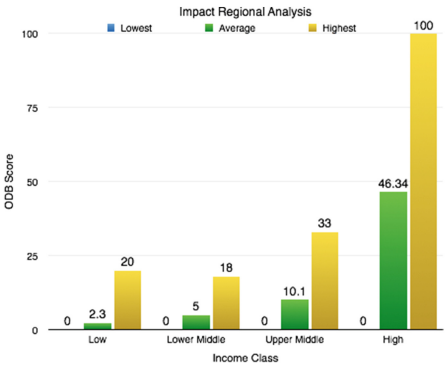


Fig. 11. Impact regional analysis

y-axis highlights the ODB sub-index score. Hence, one point on the x-axis represents the country's ODB rank and the three corresponding coloured points on the y-axis are translated into the respective sub-index scores. The graph shows that countries have higher ODB ranks due to better performing sub-indexes. It can be seen from the line graph that there are certain anomalies. For example, Chile has an overall ODB rank of 15 with the readiness sub-index of 69, the implementation sub-index of 73 and the impact sub-index of 8. Here the implementation sub-index is higher than the readiness sub-index. In fact, the impact sub-index is very low for the country as it is a part of the high income layer and has a reasonably high ODB rank as well. The open data hasn't had a noticeable impact on government efficiency, social policies and the economy mostly due to the lack of government initiatives and low entrepreneurial activity in the country.

Another similar example is that of Brazil with an overall of ODB rank of 21 with a moderate readiness sub-index, high implementation sub-index but a very low impact sub-index when compared with other counties in similar range of ODB rank. It is observed that even though the openness in Brazil's 2013 ODB results is pretty large for categories such as census, government spending, international trade etc. none of these categories actually adhere to the full open data standards hence giving it an overall moderate readiness factor. Also, it has been observed in case of Brazil that although there are a number of open data policies in place by government, the policies do not really pay much attention to the actual user perspective or overcoming the impediments of the use of open data [21]. This means that in order to improve the impact sub-index, the policies must be refined in a way that only fully open data is released benefiting the civil society at large.

We also looked at the percentage difference from lower income tier to higher income tier countries in terms of their income vs. ODB values. The World Bank classification of the four income country categories is based on GNI per capita [13]. Therefore, in Table 2 we listed these percentage differences of the average values of GNI, GNI per-capita and ODB. It can be seen that in the bottom tier jump (low to lower middle) and middle tier jump (lower middle to upper middle), the ODB value increase falls significantly behind the GNI increase. For example, at the low to lower middle

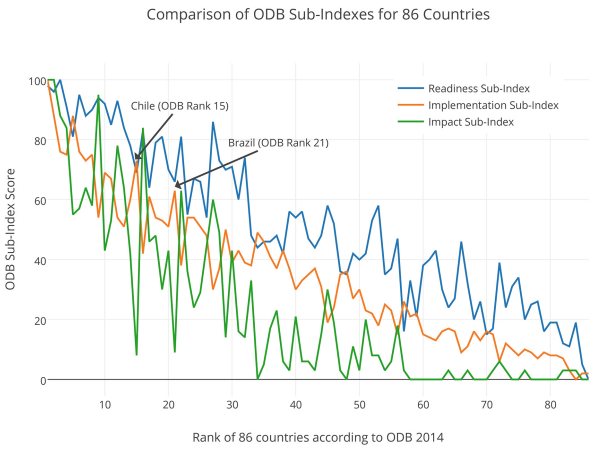


Fig. 12. ODB sub-index line graph

jump, the average ODB value percentage increase is only half of the average GNI per capita increase, and nearly 1/3 of the average GNI increase. At the top tier, however, the numbers are much more consistent. In the jump from upper middle to high class, the average ODB value percentage increase is nearly the same as that of the average GNI, and is over 60 % of the average GNI per capita increase. These numbers probably suggest that higher income countries generally put up more open data efforts, resulting in their open data status more likely matching their income level status. A more solid conclusion would require future work that looks into more details of the breadth and depth of various income and ODB parameters.

Table 2. National income and ODB values - percentage differences among income groups (11 countries without updated GNI values are not considered)

ODB sub-index	Avg. GNI (US$)	Avg. GNI per Capita (US$)	Avg. ODB values
Upper middle to high	45.17 %	78.85 %	48.18 %
Lower middle to upper middle	60.17 %	74.29 %	38.19 %
Low to lower middle	90.9 %	66.32 %	33.8 %

5 UK, US and China Comparison

After reviewing the global open data landscape, this paper compares highly ranked ODB countries like UK and USA with China in more depth. China has been taken in this analysis as it is an epitome of the developing world. Taking into consideration it's size, population and gross domestic product, it should indeed be releasing vast amounts of data, contributing to the society and making use of areas such as machine learning and business intelligence. However, government initiatives for open data pose a major

challenge to this contribution. This section analyses the current open data situation of the country and the barriers that it needs to overcome.

5.1 ODB Sub-indexes and Their Relationship

As observed previously, there is a substantial jump in the average ODB scores of countries from the upper middle income category to the countries of the high income. In the previous section, it has been concluded that this was mainly due to the implementation and impact sub-indexes. In this section, a comparison is performed between the overall ODB scores, readiness sub-index, implementation sub-index and impact sub-index of UK, US and China (Fig. 13). The US and UK can be considered examples of the high income layer, and China is an example of the upper-middle income layer. UK and US are ranked first and second respectively in the ODB rankings of 2014 [15]. The reason as to why UK and US rank so high is because of legislations in their respective countries [22]. In recent few years, these governments have launched a number of initiatives that basically target health, energy, climate, education, finance, public safety and development sectors thereby, improving open data initiatives [22].

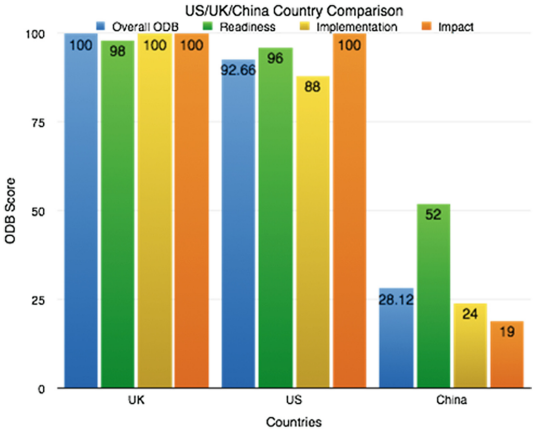


Fig. 13. UK, US, China Country comparison

China, on the other hand ranks 46th in the ODB ranking. Although its readiness sub-index is just twice as low as the other two countries, it's lagging behind in the implementation and impact sub-indexes, which is 75 % of the overall ODB weightage. Therefore, it can be observed that China needs to work on factors such as making datasets fully open data compliant in the fields of innovation, social policy and accountability, implementing strong open data legislation as well as maximizing impact in the fields of political, social, and economic importance.

5.2 National Data Portal

Both US [23] and UK [24] have created national data portals where data.gov and data.gov.uk have released 14,008 and 22,385 datasets respectively. The most common machine-readable formats for US and UK are XML and CSV, while popular non-machine readable formats for the two are HTML and PDF respectively. In addition, US also offers a significant number of datasets in zip format as compared to UK.

Unlike US and UK, China does not have a national data portal yet. However, certain open data is available through different agencies. One such example is availability of open data through National Bureau of Statistics of China (NBSC), which offers both a Chinese and an English version. The Chinese version is organized into monthly, quarterly, annual, regional, international and census data. The English version only consists of four categories i.e. monthly, quarterly, annual and regional data. All data from the English version is in machine-readable format. However, the same doesn't apply for the Chinese version, as it is only available in HTML format.

5.3 Dataset Example – Air Quality Index

In this section, we look at an example dataset common in all three countries, the air quality metrics, which measures the air pollutants level in daily air quality. The data is from the US Environment Protection Agency [25], U.K. Department for Environment Food and Rural Affairs [26], and the Chinese Ministry of Environmental Protection [27]. In U.S. and U.K. the metrics is reported by Air Quality Index (AQI), while in China, prior to 2012 the metrics was reported by Air Pollution Index (API). China switched to AQI in 2012 [27]. From the datasets it is clear that for UK and US AQI data is consistently available all regions and major cities since year 2000.

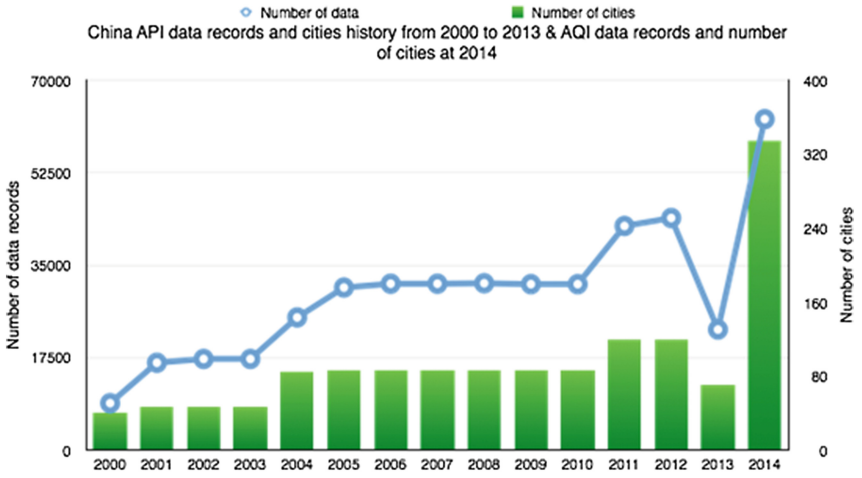


Fig. 14. China API and AQI data records

China's air quality metrics, on the other hand, has been reported with API since 2000. Figure 14 shows that in 2013 the number of reporting cities as well as the number of API datasets reduced dramatically due to the transition process from API to AQI. However, there is an obvious improvement afterwards as the number of cities with available AQI data increased to 335 in 2014 from 120 in 2012 with API data.

According to the ODB measurement methodology and weightage mechanism [15], we have calculated the ODB score for AQI data for the three countries. Our findings

were, on a full score of 100, UK obtains 95 with the only limitation of no linked data URLs. U.S. and China receive scores of 80 and 50 respectively; the main reasons that pull China's score down are related to machine-readability and ability to download data. Since the Chinese dataset is already provided in HTML, technically it is not difficult to incorporate both of these aspects. Another reason for reduced score is an explicit link for open-license, which should be even easier to address.

6 Conclusion

We presented a heat-map analysis of the global open data landscape based on the ODB ranking system. We looked at the overall ODB indexes and its associated readiness, implementation and impact sub-indexes, for countries of the low, lower-middle, upper-middle and high income-levels, respectively. Our results show that in many countries the three sub-indexes do not exhibit dependency on each other. The impact sub-index is found to be most often the weakest part of the three, and in quite a few cases extremely low compared to the readiness and implementation indexes. This observation shows that, on one hand, governments around the world are establishing more and more open data initiatives and citizens are engaging in an increasing number of open data activities; on the other hand, tangible political, social and economic benefits from open data remain to be seen. This may be because we are still in the early stage of the entire open data life cycle where harvests are yet to be reaped. However, the extremely large gaps between the impact and other two sub-indexes in some countries may warrant a thoughtful review of the existing open data initiatives to more effectively align the investments with the expected results of open data.

In addition to the global open data perspective, we also provide a comparison of China with leading open data advocates, UK and US. We found that although China is lagging behind the two other countries in the three sub-indexes, gap in the implementation and impact sub-index is much larger than the gap in readiness sub-index. This shows the Chinese government has made good progress in facilitating open data initiatives from the policy and regulatory front, but more needs to be done especially on how to put those policies into execution, which is crucial for a positive impact of open data. Follow-up example of air quality data further confirmed China's clear progress in making data ready, but not yet providing data optimal for implementation. In summary, we believe that a number of natural steps can be taken to boost China's open data status to its next level, e.g., establishing a national and regional data portals will facilitate interested parties to find the right data; making available more machine-readable data will dramatically improve its usability and value.

References

1. Open Knowledge. What is Open? (2015). https://okfn.org/opendata/. Accessed 12 Feb 2015
2. Goldstein, B., Dyson, L.: Beyond Transparency: Open Data and the Future of Civic Innovation, 1st edn. Code for America Press, San Francisco (2013)

3. Huijboom, N., Van den Broek, T.: Open data: an international comparison of strategies. Eur. J. ePractice 12, March/April 2011
4. Heimstädt, M., Saunderson, F., Heath, T.: Conceptualizing Open Data ecosystems: A timeline analysis of Open Data development in the UK, Discussion Paper, School of Business & Economics: Management, No. 2014/12
5. Kassen, M.: A promising phenomenon of open data. Gov. Inf. Q. **30**(4), 508–513 (2013)
6. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits: adoption barriers and myths of open data and open government. Inf. Syst. Manage. **29**(4), 258–268 (2012)
7. Zuiderwijk, A., Janssen, M.: Open data policies, their implementation and impact: a framework for comparison. Gov. Inf. Q. **31**(1), 17–29 (2014)
8. Hartmann, P.M., Zaki, M., Feldmann, N., Neely, A.: Big Data for Big Business? Working Paper, Cambridge Service Alliance, University of Cambridge (2015). http://www.cambridgeservicealliance.org/. Accessed Feb 2015
9. Magalhaes, G., Manley, L.: Business models for open government data. In: Proceedings of the 8th Annual International Conference on Theory and Practices of Electronic Governance (ICEGOV), Guimarães, Portugal, Oct 2014
10. WWW Foundation (2014). http://opendataresearch.org/sites/default/files/posts/Common%20Assessment%20Workshop%20Report.pdf. Accessed 20 Feb 2015
11. Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., Schraefel, M.C.: Linked open government data: lessons from Data.gov.uk. IEEE Intell. Syst. **27**(3), 16–24 (2012)
12. World Urbanization Prospects - The 2014 Revision (2014). http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf. Accessed 5 Mar 2015
13. The World Bank. Data (2015). http://data.worldbank.org/about/country-and-lending-groups. Accessed 21 Mar 2015
14. President Xi's Requirements on Jing-jin-ji Coordinated Development Plan (2015). http://news.xinhuanet.com/politics/2014-02/27/c_119538131.htm. Accessed 19 Jun 2015
15. Open Data Barometer, 2nd edn. (2015). http://www.opendatabarometer.org/report/about/method.html. Accessed 5 Mar 2015
16. The World Bank – Open Government Data Working Group (2013). http://data.worldbank.org/sites/default/files/1/od_readiness_-_revised_v2.pdf. Accessed 20 Feb 2015
17. Open Knowledge Foundation – Open Data Index (2014). https://index.okfn.org/country. Accessed 19 April 2015
18. Capgemini Consulting – The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data (2013). http://ebooks.capgemini-consulting.com/The-Open-Data-Economy/#/4/. Accessed 19 April 2015
19. Susha, I., Zuiderwijk, A., Janssen, M., Gronlund, A.: Benchmarks for evaluating the progress of open data adoption: usage, limitations, and lessons learned. Soc. Sci. Comput. Rev. (2014). doi:10.1177/0894439314560852
20. ePSI Platform – The PSI Scoreboard (2015). http://www.epsiplatform.eu/content/european-psi-scoreboard. Accessed 19 April 2015
21. Matheus, R., Ribeiro, M.M.: Open data in legislative: The case of Sao Paulo (2014). http://www.opendataresearch.org/content/2014/665/open-data-legislature-case-s%C3%A3o-paulo-city-council. Accessed 25 Mar 2015
22. The White House. Open Data Initiative (2013). https://www.whitehouse.gov/blog/2014/05/09/continued-progress-and-plans-open-government-data. Accessed 06 Apr 2015
23. Data.Gov (2015). http://www.data.gov/. Accessed 2 Feb 2015
24. Data.Gov.UK Opening Up Government (2015). http://data.gov.uk/. Accessed 2 Feb 2015

25. EPA. Air Quality Report (2014). http://www.epa.gov/airquality/airdata/ad_rep_aqi.html. Accessed 20 Apr 2015
26. Department for Environment Food & Rural Affairs (2015). http://uk-air.defra.gov.uk/air-pollution/daqi. Accessed 19 Apr 2015
27. Ministry of Environmental Protection of the People's Republic of China (2015). http://datacenter.mep.gov.cn/. Accessed 12 Apr 2015

Open Data Platforms and Their Usability: Proposing a Framework for Evaluating Citizen Intentions

Kawaljeet Kapoor(✉), Vishanth Weerakkody, and Uthayasankar Sivarajah(✉)

Brunel Business School, College of Business, Arts and Social Sciences, Brunel University London, London, UK
{Kawaljeet.Kappor,Vishanth.Weerakkody, Sankar.Sivarajah}@brunel.ac.uk

Abstract. Governments across the world are releasing public data in an effort to increase transparency of how public services are managed whilst also enticing citizens to participate in the policy decision-making processes. The channel for making open data available to citizens in the UK is the data.gov.uk platform, which brings together data relating to various public services in one searchable website. The data.gov.uk platform currently offers access to 25,500 datasets that are organized across key public service themes including health, transport, education, environment, and public spending in towns and cities. While the website reports 5,438,159 site visits as of June 2015, the average time spent on the site has been recorded at just 02:12 min per visitor. This raises questions regarding the actual use and usability of open data platforms and the extent to which they fulfill the stated outcomes of open data. In this paper, the authors examine usability issues surrounding open data platforms and propose a framework that can be used to evaluate their usability.

Keywords: Open data · Citizen · Usability · Evaluation · Public services

1 Introduction

The push for making public services data available to the community started around mid 2000s with the European Union directive encouraging greater realisation of the economic value of public data through its reuse in 2003. This directive, combined with the advancement of Internet and associated ICT tools facilitating data analytics [29], has paved way for governments to open up data to the community. Conventionally, government departments retained public service data within their systems, with limited information being released to citizens and other stakeholders (businesses, charitable organizations, and NGO communities). However, since the last decade, the spread of digital governance and associated norms such as responsiveness, accessibility and efficiency of public services, transparency and accountability, have motivated governments to exploit the potential of wider distribution and use of such data [28]. One of the first countries to mandate the use of open data was the United States. The Presidential Open Government Directive in December 2009 required the use of open

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 261–271, 2015.
DOI: 10.1007/978-3-319-25013-7_21

formats by all federal US agencies. Six months later, the UK followed with their own plans for open public data, with the Prime Minister announcing the setting up of the data.gov.uk website. Subsequently, the European Commission published a Communication on Open Data in 2011, and in the same year, USA, UK, and initially six other countries were signatories to the Open Government Declaration.

The UK is considered as one of the leading countries in Europe for open data. The data.gov.uk is one of the most comprehensive open data repositories making available non-personal UK government data about public services, ranging from health, social services, education, transport to crime and other geo-environmental data. It was launched in closed-beta in September 2009 and publicly launched in January 2010 (data.gov.uk/blog/the-new-datagovuk). When the data.gov.uk website was officially launched in January 2010, ordnance survey data which provides information on geographical locations was one of the key datasets that was opened up as part of the project [3]. Subsequently, in June 2010, the Treasury released the Combined Online Information System (COINS), which operates as the UK Government's central accounting system detailing the spending of all government departments and their major spending programmes [38]. As of June 2015, the data.gov.uk website contains over 25500 datasets. The data can be used by individuals, businesses and other stakeholders under the conditions that the copyright and the source of the data is acknowledged by including an attribution statement specified by data.gov.uk.

One of the motivations of the open data movement has been to make more data easily accessible to diverse stakeholders with a view of enticing them to participate and contribute to the public policy-making space. For example, it is anticipated that researchers, think-tanks, entrepreneurs, businesses leaders, representatives of public services, NGOs, charities, community groups and citizens, at large, will use the open data to contribute to the policy decision making process, particularly across local governments. Indeed, this has encouraged these stakeholders to take an active interest in the way services are currently being delivered, and has stimulated thinking around how to improve services. Although countries such as the UK and US have taken proactive steps to improve the availability and ease of use (through machine-readability and technical standards), there still remain several barriers to accessibility and usability of open data. Moreover, the actual use of open data is cumbersome and stakeholders have to do it themself. In order to fully exploit the potential of open data, users will usually require a certain level of applied skills. The fact that there is no existing easy-to-use, proven solution which can help citizens exploit the open data for decision making regarding their own lives, or contribute to the wider public policy making debate, does not help. Such issues are further compounded by the generic nature of open data repositories such as www.data.gov.uk and www.epsiplatform.eu, and thus their relevance and direct interest to citizens, in particular.

The motivation for this paper lies in the reasoning that although the availability of open data offers many opportunities for citizens, no research exists that questions the usability of open data platforms, particularly from a citizen's perspective. In this paper, we set out to examine and discuss some of the salient factors that influence the usability of open data by citizens and propose a conceptual framework to encapsulate these factors. In order to do this, we first review the benefits and challenges of open data followed by examining the role of open data platforms, and motivations for using such

platforms. Thereafter, we identify potential measures for evaluating the usability of open data platforms and propose a framework to capture these. We conclude the paper by offering a discussion to synthesise the main arguments presented in the paper, identifying the main limitation, and pointing at future research directions.

2 Open Data: An Overview

The goal of Open Data initiatives has been to open all non-personal and non-commercial data, especially data collected and processed by government organizations [1]. It can be seen as a movement very similar to the Open Source or Open Access phenomenon. In the course of this trend, public sector organisations have started making governmental data available on web portals, as web services so that the public have access to these data at a single point of access to official datasets. The increase in availability of open data initiatives has been seen as mainly due to the growing pressure imposed by governments on all kinds of public organisations to release their raw data [18]. The key motivators to encourage public organisations for publishing data revolves around government's perception that the open access to publicly-funded data provides: (a) greater economic returns from public investment [11], (b) provides policy-makers with data needed to address complex problems [7], (c) generate wealth through the downstream use of outputs [18], and (d) help involve citizens in analysing large quantities of datasets [30]. In general, the overarching arguments for stimulating open data are highlighted as the increase in political economic growth and the contribution to public values (i.e. transparency and accountability).

2.1 Benefits and Challenges in Using and Accessing of Open Data

Many scholars believe that Open data can be a valuable resource of information if published in a useful manner (e.g. [1, 2]). Some of the key benefits and challenges identified in the extant literature (e.g. [10, 11, 19, 32] are synthesised and presented in Tables 1 and 2, respectively. These tables do not provide an extensive list of benefits and challenges of open data use, but highlight the prominent opportunities and concerns discussed in the literature.

Although open data can potentially provide many benefits, its use also comes with a number of challenges. Some of the key challenges identified in the extant literature are presented in the following table.

2.2 Use of Open Data Platforms

The main purpose of open data platforms has been to promote access to government data and encourage development of creative tools and applications to engage and serve the wider community [22]. In doing so, enabling civic engagement by providing opportunity for citizens, public sector organisations, businesses and independent developers to use systematically-updated stream of open data is being encouraged.

Table 1. Benefits of open data

Benefits	Description	References
Increased transparency and accountability	Making government data transparent should increase public trust in government and civil servants and also allow citizens to hold the government officials accountable	[11, 19, 32]
Economic growth	Opening government data is believed to bring a range of economic benefits such as encouraging the marketplace to develop products and services, which increase productivity, offer employment, and bring revenue back to government in the form of taxation revenue	[1, 19]
Societal benefits	Potential to allow citizens to interact with government in a more informed and interactive manner	[32]
Cost reduction and efficiencies	Sourcing data is often significantly costly in both time and money for organisations. Opening government data can significantly reduce the costs associated with acquiring data	[33]
Improved data quality	Leads to improved data quality via crowdsourcing of corrections or by filling gaps in data	[11]
Simulating innovation	Opening government data encourages developers and the general public to explore and play with new data that might lead to development of innovative solutions	[19, 32]

The governments perceive that making this data available on the web would lead to more transparency, participation, and innovation throughout society [10, 19]. Often open data platforms publish datasets covering a wide range of domains, from environmental data over employment statistics to the budgets of municipalities. Publishers of these datasets can be individual government agencies or providers of larger repositories that collect public datasets and make them available in a centralized and possibly standardized way. Governments and publishers of open data expect the users to exploit these data in many ways as possible for the benefit of the society [12]. For example, general public (non-technical users) may use it simply to analyse trends over time from one policy area, or to compare how different parts of government go about their work. On the other hand, technical users such as software developers are encouraged to create useful applications out of the raw data files, which can then be used by everyone benefitting the wider society.

In terms of the process to find the available open data, end-users of these platforms (i.e. citizens, businesses) who wish to access and use Open Data need to first identify relevant datasets manually or by visiting a central repository/platform (e.g. data.gov.uk). In the case of finding datasets manually, this includes finding organizations or agencies that publish open datasets on platforms that provide a central and responsive entry point where users can search for data. If a single dataset can be found, that

Table 2. Challenges of open data

Challenges	Description	References
Increasing public interest	Challenge of raising the capacity and awareness of civil servants, citizens and the private sector on their rights to access and re-use public data initiatives	[32, 40]
Cost of opening up data	Time and resource costs are seen as obstacles for government departments in opening their data, especially as they were often experienced as upfront costs	[11, 32]
Data ownership risks and legality concerns	In the past, if councils or government departments contracted a third party to gather data for them, or purchased data directly from the third party, they often licensed the data and did not own the intellectual property rights, and thus could not directly release it under open copyright	[18, 41]
Uncertainty about data stream continuity	If a user is not positive that a data stream will be maintained in the future, this creates uncertainty around any project using that data stream. This reduces the chances that an organisation or individual will be willing to invest the time and resources into a product or application that uses this data	[18, 32]
Data quality concerns	Government departments may be reluctant to release data that they see as low quality. Some agencies are worried about the potential liabilities of releasing their data concerning information accuracy, up-to-dateness etc.	[10, 39]
Privacy violation	Data that includes private or potentially sensitive information on citizens; there can be concerns over whether and how the data can be anonymised, what can be released, to whom and under what copyright	[9, 11]

contains all the relevant data, the user can directly extract the required information. However, it is rather unlikely to find all relevant data in a single file. The way people access and use Open Data is greatly influenced by the way the data is published [1]. Many government agencies or organizations collect large amounts of data. In its original, raw form, this data is often not very useful for end users. Therefore, many datasets are cleaned and customized before being published. While some publishers prefer the data to be in a human-readable format, others prefer a machine-readable format. Apart from accessing data from these platforms, users (e.g. organisations) are also encouraged to submit useful data that can be published to the general public. Government open data initiatives are also encouraging users in a number of ways to be involved as part of these projects dependent on their background or interest. For

example, one of the challenges is making existing data come to life, and users are encouraged to combine and reorganise existing data to offer new insights resulting in useful visualisations of these data [12].

3 Developing Measures for Evaluating the Acceptance of Open Data Platforms

Websites such as data.gov.uk make it easy for citizens to access governmental data and other offered services whilst increasing citizens' potential of contributing to democratic processes [14]. According to Wangpipatwong et al. [37], citizen use of such websites substantially reduces the management and operational costs for the government. This study aims to empirically investigate the use of the aforementioned open data website from a citizen's perspective. A suitable mix of measures will be borrowed from the available innovation adoption models to evaluate the citizens' continued use intentions of such websites. This will be undertaken by gathering the opinions of those who already have the experience of using data.gov.uk along a set of measures identified from the literature.

Available literature shows that very few studies have attempted to empirically evaluate the performance of open data websites. There are, however, evidences of other studies using different measures of innovation adoption to investigate the performance of different websites. For instance, Wangpipatwong et al. [37] use the Technology Acceptance Model (TAM) alongside self-efficacy as an added measure, to evaluate the use of an e-government website. Fang and Holsapple [15] focus on the navigation structure of a website and their impact on the usability of that website by using factors defining its usability. Wang and Senecal [36] used ease of use, speed, and interactivity to measure the usability of a website and its subsequent impact on user attitudes and intentions.

The literature is rich with theoretical models, mostly developed from the psychology and sociology theories, which assist in analysing the acceptance of a service or a product [34, 35]. Some of the most used models come from the following theories: Diffusion of Innovations theory (DOI) by Rogers [25], Theory of Reasoned action (TRA) by Fishbein and Ajzen [16], Theory of Planned Behaviour (TPB) by Ajzen [4] and Ajzen and Fishbein [5], Technology Acceptance Model (TAM) by Davis [13], Decomposed Theory of Planned Behavior, Extended Technology Acceptance Model, and Unified Theory of Acceptance and Use of Technology by Venkatesh et al. [35]. The DOI theory is regarded as a principal theoretical perspective on technology adoption offering a conceptual framework for discussing adoption at a global level. Rogers [27] has synthesized sixty years of innovation-adoption research in developing this theory. His DOI model has been well received in the world of innovative solutions, and it is one of the most used theories in the field of innovation diffusion [20, 31].

Rogers [27] identified the following five attributes as the perceived attributes of innovations within DOI – relative advantage, compatibility, complexity, trialability, and observability. It can be easily observed from the attributes used in the aforementioned models that the TPB model is an extension of the TRA model, and the decomposed TPB model shares similarities with TAM. Fishbein and Ajzen [16] incorporated attitudes, subjective norms and behavioural intention in their TRA model.

TAM is also regarded as an adaptation of the TRA model and the TAM model also shares two attributes with the DOI model (relative advantage/perceived usefulness and complexity). Davis [13] identified perceived usefulness and ease of use alongside the effects of attitude on intention in their TAM model, as the factors influencing the acceptance of a technology. Giving due consideration to all of these innovation adoption models, the following attributes were shortlisted depending upon their relevance to the case of open data website being covered within this study (Fig. 1): perceived usefulness, compatibility, ease of use, result demonstrability, trust, risk, social approval, visibility, and behavioural intentions.

Perceived usefulness, also referred to as the relative advantage, will help assess if the information available on the website is relatively better across multiple aspects in comparison to the same data that a citizen can access via other physical offices and platforms. In measuring the advantages of a new service, users tend to evaluate the pluses and minuses of using that service. This characteristic is known to determine the ultimate rate of most innovation adoptions in the long run [24]. In terms of compatibility, the website will be assessed for the type of information it offers to the citizens with respect to the type of information the citizens are interested in, or are expecting to, access using such open data platforms. Rogers [27] describes compatibility to be the degree with which the introduced innovation manifests itself as being consistent with users' past experiences, present values, and their future needs.

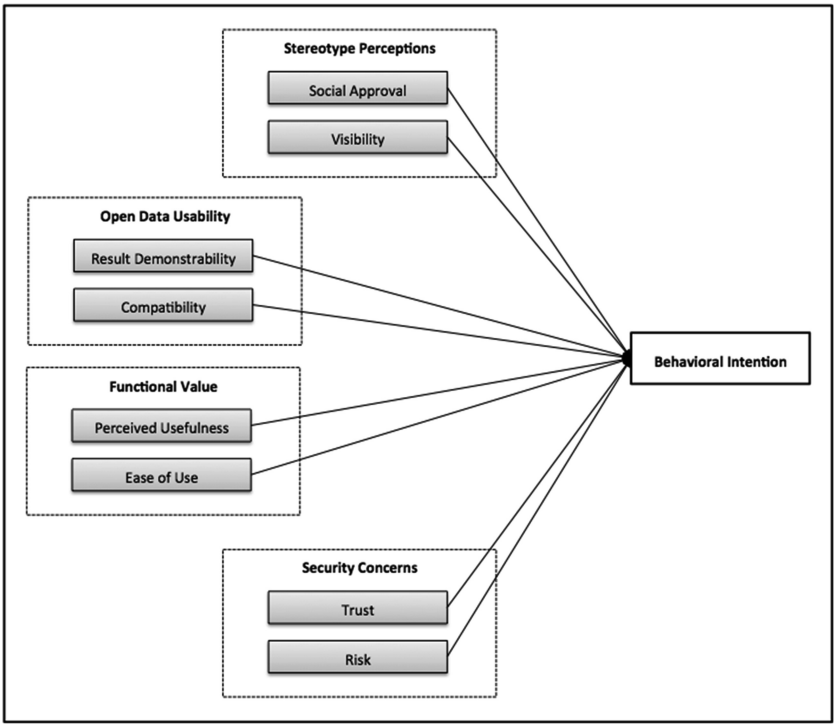


Fig. 1. Proposed framework for open data websites

Users' knowhow of a service tends to dictate their perception of the level of effort involved in using that service. The less complex a service is to use/operate, the more easily it is accepted. The ease of using the data.gov.uk website will be evaluated from a perspective of optimizing user experience; the design of the interface, time required to look up for the desired information, understandability of different features/tabs offered within the website, and any other navigation complexities will be explored using this attribute (page layout, scrolling and paging, text appearances, links, search and so on). Result demonstrability will be measured along users' satisfaction of the quality of information that they can retrieve using the open data website. The trust factor will be used to measure the extent to which the users trust the validity of the information that they are accessing using such open data websites. A user's uncertainty about the quality of information being gathered can potentially lead to anxiety, which can come in the way of their adoption decision. The expected social or economic loss caused from using a new system constitutes perceived risk [21, 26]. In this context, the risk factor will be used to measure users' apprehensions of inputting personal information onto such websites, and also their confidence in using the information available on such websites.

The members of a social system generally tend to display a sense of belonging by being a part of the activities that are regarded as a norm within their social system [23]. Social interaction and information exchange can play critical roles in promoting the use of a new system, in effect, motivating individuals to accept that system [6]. Social approval is a construct that will help measure approval from a user's social circle (friends, families, and peers) regarding the use of open data platforms and their authenticity. Visibility of a system encourages peer discussions of that system, which collectively contributes towards achieving a better acceptance rate for that system [27]. Thus, the visibility construct will be used in this study to help identify the awareness that the citizens have about the existence of such open data platforms that they can utilize to their benefit on a daily basis.

The effects/influences of perceived usefulness, compatibility, ease of use, result demonstrability, trust, risk, social approval, and visibility will then be individually measured across users' behavioural intentions. Behavioural intention is perceived as an instinctive probability that a user relates directly with the possibility of a particular behavioural option being chosen [8]. Some models of innovation adoption and diffusion (TRA and TPB) acknowledge this attribute to be the best immediate predictor of the acceptance of a given service/product [23]. These eight characteristics are expected to positively and significantly impact users' intentions towards the usage of open data platforms. The positive or negative correlations that will surface post the empirical evaluations will then be logically reasoned and analysed for their significance in influencing users' intentions towards using the data.gov.uk website.

4 Discussion and Conclusions

This paper focuses specific interest on open data platforms to establish an understanding of its usability from a citizen perspective. In the UK, data.gov.uk is the functional open data website available to be accessed by the public. The UK

government has introduced an action plan for a smarter and more efficient government to open the government and promote transparency in empowering citizens and their roles in the civic society; their focus is on redefining the relationship between the frontline and the centre to better manage finances via efficient delivery [33].

In terms of theoretical implications, this study broadly touches upon the streams of open government, ICT literature, and digital governance. The framework presented in this paper has been developed from well-established and most used theories in innovation adoption. This framework can be applied across different open data domains to assess the impact of transparent governance on empowering and encouraging citizen engagement in open government data initiatives. Other researchers can use this framework to build upon, as required, to run in-depth analyses of specific aspects (such as trust in available information, level of risk involved in utilizing that information, and so on) of open data and its usability across different contexts.

In terms of practical implications, a significant milestone of this research is the development of a framework that would enable a reliable assessment of the open data platforms. A 2015 report from a four million project funded by the European commission revealed that one of the issues of such open data platforms is that most data owners in the public sector hold a poor understanding of the relative benefits of different data formats [17]. Most owners tend to adopt a path of least resistance and publish the data in its original format, ignoring the potential of making the data available in more reader-friendly capable formats. Nevertheless, such open data holds numerous practical implications for international data standard forums [17]. The framework will assist policymakers, publishers of open data, IT practitioners (application developers), and other proactive citizens in analyzing the usability of open data platforms. This framework will help evaluate the perceived usefulness of readily available open data, whilst measuring its compatibility with user needs. The stakeholders will also be able to assess the quality of information available in these websites across the risk and trust perceptions of the involved users. Other issues such as user friendliness and usefulness will also be measurable across the aspects of ease of use and result demonstrability. Overall, the constructs put together in the framework will help practitioners to summarize the effectiveness of the specific open data platforms being evaluated, to be eventually bettered for future use by the citizens.

This study suffers from the limitation of no empirical evidence supporting the validity of the proposed framework. This is an ongoing research, and having established a framework for evaluating open data, our future research will be focused on empirically assessing the validity of this framework in the context of a UK open data website: data.gov.uk. Exclusive scenarios will be defined prior to the empirical evaluations; for instance, the users will be directed to explore specific categories (housing, environment, taxes, health, and so on) in the targeted website to assess the retrieved results against the framework proposed within this study. This will help analyze and identify problem areas, if any, to be marked for improvement by the publishers of such open data.

References

1. Borzacchiello, M.T., Craglia, M.: The impact on innovation of open access to spatial environmental information: a research strategy. Int. J. Technol. Manage. **60**(1–2), 114–129 (2012). doi:10.1504/ijtm.2012.049109
2. Braunschweig, K., Eberius, J., Thiele, M., Lehner, W.: The State of Open Data Limits of Current Open Data Platforms (2012)
3. BBC News. Ordnance Survey offers free data access. http://news.bbc.co.uk/1/hi/8597779.stm (2009). Accessed 10 Jun 2015
4. Ajzen, I.: From intentions to actions: a theory of planned behaviour. In: Kuhl, J., Beckman, J. (eds.) Action-Control: from Cognition to Behavior, pp. 11–39. Springer, Heidelberg (1985)
5. Ajzen, I., Fishbein, M.: Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs (1980)
6. Bandura, A.: Social Foundations of Thoughts and Action: a Social Cognitive Theory. Prentice-Hall, Englewood Cliffs (1986)
7. Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Wouters, P.: An international framework to promote access to data. Science **303**, 1777–1778 (2004)
8. Chiu, R.K.: Ethical judgment and whistleblowing intention: examining the moderating role of locus of control. J. Bus. Ethics **43**(1–2), 65–74 (2003)
9. Choenni, S., van Dijk, J., Leeuw, F.: Preserving privacy whilst integrating data: applied to criminal justice. Inf. Polity **15**, 125–138 (2010)
10. Conradie, P., Choenni, S.: On the barriers for local government releasing open data. Gov. Inf. Q. **31**, S10–S17 (2014)
11. Cranefield, J., Robertson, O., Oliver, G.: Value in the mash: exploring the benefits, barriers and enablers of open data apps. In: Proceedings of the European Conference on Information Systems (ECIS) 2014, Tel Aviv, Israel, 9–11 June 2014. ISBN 978-0-9915567-0-0
12. Data.gov.uk.: About data.gov.uk. http://data.gov.uk/about (2015)
13. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**, 319–340 (1989)
14. Fang, Z.: e-government in digital era: concept, practice, and development. Int. J. Comput. Internet Manage. **10**(2), 1–22 (2002)
15. Fang, X., Holsapple, C.W.: An empirical study of web site navigation structures' impacts on web site usability. Decis. Support Syst. **43**(2), 476–491 (2007)
16. Fishbein, M., Ajzen, I.: Belief, attitude, intention, and behavior: an introduction to theory and research. Addison-Wesley, Reading (1975)
17. Glidden, J.: The Citadel Reveals Open Data Findings. https://opensource.com/government/15/6/citadel-open-government-data-results (2015)
18. Janssen, K.: Open government data and the right to information: opportunities and obstacles. J. Community Inform. **8**(2) (2012). http://www.ci-journal.net/index.php/ciej/article/view/952/954
19. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manage. **29**(4), 258–268 (2012)
20. Kapoor, K., Dwivedi, Y.K., Williams, M.D.: Role of innovation attributes in explaining the adoption intention for the interbank mobile payment service in an indian context. In: Dwivedi, Y.K., Henriksen, H.Z., Wastell, D., De', R. (eds.) TDIT 2013. IFIP AICT, vol. 402, pp. 203–220. Springer, Heidelberg (2013)
21. Labay, D.G., Kinnear, T.C.: Exploring the consumer decision process in the adoption of solar energy systems. J. Consum. Res. **8**(3), 271–278 (1981)

22. Martín, A.S., de Rosario, A.H., Pérez, C.C.: Open government data: a european perspective. In: Information and Communication Technologies in Public Administration: Innovations from Developed Countries, vol. 195 (2015)
23. Ozaki, R.: Adopting sustainable innovation: what makes consumers sign up to green electricity? Bus. Strategy Environ. **20**(1), 1–17 (2011)
24. Pannell, D.J., Marshall, G.R., Barr, N., Curtis, A., Vanclay, F., Wilkinson, R.: Understanding and promoting adoption of conservation practices by rural landholders. Aust. J. Exp. Agric. **46**(11), 1407–1424 (2006)
25. Rogers, E.M.: Diffusion of Innovations. The Free Press, Glencoe (1962)
26. Rogers, E.M., Shoemaker, F.F.: Communication of Innovations. Free Press, New York (1971)
27. Rogers, E.M.: Diffusion of Innovations, 5th edn. Free Press, New York (2003)
28. Sivarajah, U., Irani, Z., Weerakkody, V.: Evaluating the use and impact of Web 2.0 technologies in local government. Gov. Inf. Q. (2015). ISSN 0740-624X, http://dx.doi.org/10.1016/j.giq.2015.06.004
29. Sivarajah, U., Irani, Z., Jones, S.: Application of Web 2.0 technologies in e-government: a United Kingdom case study. In: 47th Hawaii International Conference on System Sciences (HICSS), pp. 2221–2230. IEEE (2014)
30. Surowiecki, J.: The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business Economies, Societies and Nations. Doubleday, New York (2004)
31. Tornatzky, L.G., Klein, K.J.: Innovation characteristics and innovation adoption-implementation: a meta-analysis of findings. IEEE Trans. Eng. Manage. **29**(1), 28–43 (1982)
32. Ubaldi, B.: Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives, OECD Working Papers on Public Governance, No. 22, OECD Publishing, Paris (2013). doi:http://dx.doi.org/10.1787/5k46bj4f03s7-en
33. Veljković, N., Bogdanović-Dinić, S., Stoimenov, L.: Benchmarking open government: an open data perspective. Gov. Inf. Q. **31**(2), 278–290 (2014)
34. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**, 425–478 (2003)
35. Venkatesh, V., Thong, J., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**, 157–178 (2012)
36. Wang, J., Senecal, S.: Measuring perceived website usability. J. Internet Commerce **6**(4), 97–112 (2007)
37. Wangpipatwong, S., Chutimaskul, W., Papasratorn, B.: Understanding citizen's continuance intention to use e-government website: a composite view of technology acceptance model and computer self-efficacy. Electron. J. e-Gov. **6**(1), 55–64 (2008)
38. Wilcox, J.: Government drops first set of COINS. PublicTechnology.net (2010). http://legacy.publictechnology.net/sector/central-gov/govt-drops-first-set-coins
39. Zhang, J., Dawes, S.S., Sarkis, J.: Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. J. Enterp. Inf. Manage. **18**, 548–567 (2005)
40. Zuiderwijk, A., Janssen, M.: Open data policies, their implementation and impact: a framework for comparison. Gov. Inf. Q. **31**(1), 17–29 (2014)
41. Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.S.: Socio-technical impediments of open data. Electron. J. e-Gov. **10**(2), 156–172 (2012)

E-Business, E-Services and E-Society

Enabling Flexible IT Services by Crowdsourcing: A Method for Estimating Crowdsourcing Participants

Yiwei Gong(✉)

School of Information Management,
Wuhan University, Wuhan 430072, Hubei, People's Republic of China
yiweigong@whu.edu.cn

Abstract. Crowdsourcing has become an increasingly attractive practice for companies to execute business processes in open contexts with on-demand workforce and higher level of flexibility. One of the challenges is the identification of the best-fit crowdsourcing participant from a group of online candidates. This paper presents a method of AHP-TOPSIS based on Grey Relation Analysis for estimating participants of a crowdsourcing task based on their online profiles and proposals. This method is tested by an experiment on a dataset of 348 completed IT service crowdsourcing tasks. An analysis on the matching between the test result and the actual selection result reveals the accuracy and efficiency of this method. Companies can use this method to facilitate the quality control at the beginning of crowdsourcing and keeps the selection of participants easy. This paper contributes to the design of a software agent for crowdsourcing platforms to automatically rank the participants of a task.

Keywords: Crowdsourcing · Flexibility · AHP · TOPSIS · Grey relation analysis

1 Introduction

The concept of crowdsourcing is first coined in 2006 and simply means outsourcing certain tasks and problem formulations to an undefined (and generally large) network of people in the form of open calls [1]. With today's development of Internet and mobile technologies, and the explosion of social media, companies are able to have a better engagement of distributed crowds of individuals for their innovation and problem-solving needs [2]. As a result, an increasing number of companies, ranging from small startups to those listed in Fortune 500, are trying to make use of crowdsourcing to access knowledge and skills that previously unavailable to them and to solve parts of business processes formerly executed in-house [3]. In this way, companies can have a more flexible workforce and higher knowledge absorptive capacity and business processes can be adapted on-demand, which results in higher level of process flexibility [4].

Crowdsourcing can be considered as an online and distributed problem-solving model [5] and suggests that engaging crowds can help companies develop solutions to a variety of business challenges. As the business challenges and tasks vary, so do the

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 275–286, 2015.
DOI: 10.1007/978-3-319-25013-7_22

knowledge and skills that crowdsourcing participants have. Unlike simple and low-priced tasks that commonly require general skills, IT service tasks are knowledge intensive and require crowdsourcing participants with special skills and knowledge. This makes the identification of suitable participants a challenge. Matching skills to tasks often relays on sophisticated online crowdsourcing platforms to manage distributed workers and support task providers [2]. However, prior research on crowdsourcing focuses on its business models [6, 7] or brand-related and marketing alike activities [8, 9], taken the online crowdsourcing platforms and their functionalities as given [10]. The state of the art in crowdsourcing practice still lacks approaches for automated estimating participants considering their skills and knowledge [11].

In this paper, we propose a method of AHP-TOPSIS based on Grey Relation Analysis to help companies estimate and identify the best-fit participant for their IT service crowdsourcing tasks. The method is tested on a dataset of 348 completed crowdsourcing tasks in the IT service domain. A post-hoc analysis on the matching between the estimation result and the actual decision made by task providers reveals the accuracy and efficiency of this method.

2 Research Context

2.1 Forms of Crowdsourcing

The way of using crowdsourcing to abstain flexible workforce for business process execution is similar to cloud computing where computing capacity is provided on demand [11, 12]. A typical form of crowdsourcing is publishing the request for proposals through an online marketplace platform with the details of the needed service and its expected duration and (a range of) cost. Then potential participants bid on the task by submitting their proposals. Although many proposals would be received for a task, only the best-fit candidate will be selected to carry out the task. At the end, company can decide to accept and pay for the work, or refuse it if it does not fulfil the expectation. This marketplace form of crowdsourcing enables companies to access the vast potential of workforce with various backgrounds, while it has more flexibility and less risk than having a fixed outsourcing contract [13].

There are also other forms of crowdsourcing such as knowledge contributions (e.g. Wikipedia), rating (i.e. participants 'vote' on a given topic) and micro-task (e.g. Galaxy Zoo) where participants complete the task voluntarily. In addition, contest-based crowdsourcing is used for obtaining innovative ideas or solutions, in which all the participants make their effort and results are determined on a comparative basis and probably only the top contributor(s) would be reworded. Those forms of crowdsourcing are out of the scope of this paper, as they are less effective in providing business process flexibility.

2.2 Related Work

Estimation issue in crowdsourcing has been observed by researchers, and there are some studies for automatically estimating different submissions to find out those with

sufficient quality for a task. For example, Tarasov et al. [14] proposed a dynamic estimation of worker reliability in rating-based crowdsourcing. This approach is for detecting noisy and incompetent workers by estimating their submissions, instead of estimating the workers before the task was taken into execution.

Mechanisms for estimating crowdsourcing participants for business process execution can be found in the research of BPEL4People in social networks [15], in which a ranking method based on Hyperlink-Induced Topic Search (HITS) algorithm is provided to estimate the expertise of works in a social network. In this method, a certain skill, its expected level and the importance of a task are used as input, and the ranking result presents a list of all the suitable crowdsourcing works in a social network. The underlying concept of BPEL4People is that the flexibility of traditional SOA-based business process systems can be enhanced by enabling human-based services with very the same API used by software-based Web services. In this way, tasks would be able to match to suitable workers that are registered and active on the crowdsourcing social platform [11].

However, not all crowdsourcing platforms take the form of social networks where crowdsourcing participants work with each other in a joint task and some of them would take the role of supervisor or coordinator. Instead, many crowdsourcing platforms have a form of marketplaces, where a task is completed by only one participant exclusively. In this case, many candidates will compete for the same task by submitting their proposals and the task provider has to choose the best-fit one from them. The more candidates a task has, the more difficult for the task provider to manually estimate and identify the best-fit participant, because of information overload. It is therefore important to have an estimation method to rank all the candidates for the task provider to choose. And a reliable estimation should be a necessary functionality of online marketplace crowdsourcing platforms to support task providers for quality control and solving the problem of information overload.

2.3 AHP, TOPSIS and GRA

Selecting the best-fit crowdsourcing participant from a group of candidates is a typical Multiple Attribute Decision Making (MADM) problem [16] in which decision-making is for selecting the most appropriate one from many feasible solutions. Analytical Hierarchical Process (AHP) [17] is one of the most outstanding MADM methods, which first estimates the relationship among criteria weight and then the total value of each choice based on the obtained weight [18]. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is another outstanding MADM method, which is based on the concept that the best choice should have the shortest Euclidean distances from the positive ideal and the farthest from the negative ideal [16]. AHP and TOPSIS can be used in combination [19] where AHP is used to calculate the weights of the parameters and these weights are later used in TOPSIS.

The drawback of TOPSIS method is its linear variation of each alternatives, which cannot provide an accurate ranking between two alternatives that have the same distances to the ideal. This problem can be solved by Grey Relational Analysis (GRA) which is an effective method to solve decision making problems by generalizing

estimates under limited samples and uncertain conditions [20]. GRA is a kind of flexible measurement of curve similarity. By using GRA, the nonlinear relationship between the sequences of each alternative can be well reflected, which can compensate the inaccuracy problem of TOPSIS method.

AHP-TOPSIS based on GRA has been proved to be useful in solving MADM problems [21]. In this study, it is employed for estimating crowdsourcing participants for given tasks.

3 Estimation Method

In this section a method of AHP-TOPSIS based on GRA is proposed for estimating crowdsourcing participants. This method has the following three phases.

3.1 Phase 1: Identifying Estimation Parameters

The estimation parameters used by an algorithm-based method should be quantitative, otherwise they cannot be calculated. In addition, the data of parameters should be easy to access, otherwise the desired automation in the ranking of participants cannot be achieved. In this study, there are two underlying assumptions. The first one is that task providers and candidate participants do not know each other in actual life. This means that a task provider makes its selection decision only based on the related candidate participants' information that is available online. The second one is that task providers will insist on looking for the best-fit participant rather than shifting to other strategy like choosing the first acceptable candidate. This means that all related candidate participants should be involved in the consideration during the decision-making. In a typical marketplace crowdsourcing model, participants' online information comes from either their online profiles or the proposals that they submitted to the task. Both these two sources of information are involved in the formulation of the estimation parameters in this study. Afterwards, the parameters that cannot be quantitated has to be ignored, and the parameters that reflect the same property are merged. At the end, parameters are categorized into benefit parameters (the larger the value is, the better the solution is) and cost parameters (the smaller the value is, the better the solution is).

3.2 Phase 2: Using AHP to Calculate the Weight of Parameters

In AHP, the multi-attribute weight measurement is calculated via pair-wise comparison of the relative importance of two factors. Assuming that there are N number of decision parameters, denoted as (P_1, P_2, …, P_n), its judgment matrix would be $A = [a_n]$, in which a_n represents the relative importance of P_1 and P_2. Using the row vector average normalization proposed by Satty [17], the weight of P_i is calculated as:

$$W_i = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}} i,j = 1,2,\ldots,n.$$

3.3 Phase 3: Using GRA-Based TOPSIS to Estimate Participants

In this phase the algorithm has the following ten steps.

1. Normalizing of Initial decision matrix $X = \left(x_{ij}\right)_{m\times n}$, get the normalization matrix $Z = \left(z_{ij}\right)_{m\times n}$. $(i = 1,2,\ldots,m; j = 1,2,\ldots n$

   For benefit parameters:

$$Z_{ij} = \frac{x_{ij} - \min\limits_{i} x_{ij}}{\max\limits_{i} x_{ij} - \min\limits_{i} x_{ij}} \tag{1}$$

   For cost parameters:

$$Z_{ij} = \frac{\max\limits_{i} x_{ij} - x_{ij}}{\max\limits_{i} x_{ij} - \min\limits_{i} x_{ij}} \tag{2}$$

2. Calculating the weighted decision matrix $S = \left(s_{ij}\right)_{m\times n}$, $(i = 1,2,\ldots,m; j = 1,2,\ldots n)$.

$$S_{ij} = w_{ij} z_{ij}$$

3. Calculating the positive ideal solution S^+ and negative ideal solution S^-

$$S^+ = \left(s_1^+, s_2^+, \ldots, s_n^+\right); S^- = \left(s_1^-, s_2^-, \ldots, s_n^-\right)$$

   Where $s_j^+ = \max\limits_{i} s_{ij} = w_j; s_j^- = \min\limits_{i} s_{ij} = 0, i = 1,2,\ldots,m; j = 1,2,\ldots,n.$

4. Calculating the Euclidean distance between each solution and positive/negative ideal solution d_i^+, d_i^-

$$d_i^+ = \sqrt{\sum_{j=1}^{n}\left(s_{ij} - s_j^+\right)^2}, d_i^- = \sqrt{\sum_{j=1}^{n}\left(s_{ij} - s_j^-\right)^2}$$

where $i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$.

5. Calculating the grey relation coefficient matrix of each solution and positive/negative ideal solution L^+, L^-:

$$L^+ = \left(l_{ij}^+\right)_{m\times n}, L^- = \left(l_{ij}^-\right)_{m\times n}$$

$$\text{Where} \quad l_{ij}^+ = \frac{\min\limits_i \min\limits_j \left|s_j^+ - s_{ij}\right| + \theta \max\limits_i \max\limits_j \left|s_j^+ - s_{ij}\right|}{\left|s_j^+ - s_{ij}\right| + \theta \max\limits_i \max\limits_j \left|s_j^+ - s_{ij}\right|}$$

$$l_{ij}^- = \frac{\min\limits_i \min\limits_j \left|s_j^- - s_{ij}\right| + \theta \max\limits_i \max\limits_j \left|s_j^- - s_{ij}\right|}{\left|s_j^- - s_{ij}\right| + \theta \max\limits_i \max\limits_j \left|s_j^- - s_{ij}\right|}$$

Where $\theta \in (0, 1)$, is distinguishing coefficient, Here the value of θ is set to be 0.5.

$$\text{Simplifying the formulas}: L_{ij}^+ = \frac{\theta}{\left|z_{ij} - 1\right| + \theta}; L_{ij}^- = \frac{\theta}{\left|z_{ij} + 1\right| + \theta}$$

6. Calculating the grey relation grade of each solution and positive/negative ideal solution l_i^+, l_i^-:

$$l_i^+ = \frac{1}{n}\sum_{j=1}^{n} l_{ij}^+; l_i^- = \frac{1}{n}\sum_{j=1}^{n} l_{ij}^-$$

7. Applying nondimensionalization to d_i^+, d_i^-, l_i^+ and l_i^-

$$D_i^+ = \frac{d_i^+}{\max\limits_i d_i^+}, D_i^- = \frac{d_i^-}{\max\limits_i d_i^-}$$

$$L_i^+ = \frac{l_i^+}{\max\limits_i l_i^+}, L_i^- = \frac{l_i^-}{\max\limits_i l_i^-}$$

Where $i = 1, 2, \ldots, m$

8. Calculating the relative closeness degree:

$$P_i^+ = \frac{D_i^+}{D_i^+ + D_i^-}, U_i^+ = \frac{L_i^+}{L_i^+ + L_i^-}$$

9. Combining P_i^+ and U_i^+: $Q_i^+ = v_1 P_i^+ + v_2 U_i^+$

Where v_1 and v_2 reflect the degree of preference of decision makers, $v_1 + v_2 = 1$, and $v_1 = v_2 = 0.5$

10. Sorting solutions by the value of Q_i^+. The better Q_i^+ is, the better the solution is, and vice versa. $\max(Q_i^+)$ is the final decision.

4 Experiment

In order to evaluate the proposed method of AHP-TOPSIS based on GRA on its accuracy and efficiency in estimating crowdsourcing participants, an experiment was carried out on data from a popular Chinese crowdsourcing marketplace platform, Epweike (http://www.epweike.com/). This experiment used the proposed method to estimate crowdsourcing participants of certain tasks, and then the estimation result was compared with the actual selection decision made by the task provider. This comparison allows for an analysis on the accuracy of the method and the impact of the number of candidates on the actual decision-making which reflects the efficiency of this method.

4.1 Dataset

In this experiment, a dataset of 348 valid and completed tasks between 2010 and 2015 for IT services crowdsourcing is used. The content of those task includes software/mobile application development, website construction, database and system design, server maintenance, etc. In those tasks the number of openly visible proposals is more than 3. Tasks that have less than 4 proposals are ignored, because the GRA-based TOPSIS algorithm does not return a meaningful ranking result when the number of candidates is less than 4.

4.2 Approach

Phase 1: Identifying Estimation Parameters. The following 10 parameters of crowdsourcing participants are identified for the estimation, and the data of these parameters can be accessed from the website openly (Table 1).

These parameters describe the information of either the participant itself or the proposal it provided. Among those parameters, P1, P2, P3, P4, P5, P9 and P10 are benefit parameters; while P6, P7 and P8 are cost parameters.

Table 1. The description of each parameter

Parameters	Description
P1	The total volume of the participant in the history
P2	The praise rate accumulates from the assessments made by task providers for each task completed by the participant
P3	The number of biddings that the participant participates
P4	The degree of matching between the required skills for the task and the skills that the participant has
P5	The website's evaluation of the participant on its intelligence, authenticity, trusted transactions and public praise
P6	The website's overall evaluation of the participant
P7	The price proposed by the participant
P8	The task duration proposed by the participant
P9	The number of visitors of participant's homepage
P10	VIP level

Phase 2: Using AHP to Calculate the Weight of Parameters. By AHP, the weight of the 10 parameters is calculated:

$$w_i = (0.140911227, 0.194917168, 0.051377508, 0.041047892, 0.038491227, 0.030836462, 0.255857835, 0.209338244, 0.016223402, 0.020999035).$$

Phase 3: Using GRA-based TOPSIS to Estimate Participants. For space reason, only the calculation of Task1 is given as an example. This task has 23 candidate participants and 18 openly available proposals. The other 5 proposals are closed and only visible to the task provider, and therefore those 5 participants are not taken into account in the experiment. The best one is identified from the rest 18 candidates by using the GRA-based TOPSIS method. Through the estimation, the final value of Q_i^+ is presented in the following Table 2.

By descending the order of Q_i^+, the sort of participants of Task 1 can be get.

According to the result presented in Table 3, the best candidate for Task 1 is Participant 20. The values of each of its parameters are:

$$P_{20} = (17000, 100\ \%, 4, 4, 40, 5, 800, 5, 12624, 1)$$

But the subjective choice made by the task provider is Participant 1, and the values of its parameters are:

$$P_1 = (9200, 100\ \%, 4, 3, 39, 6, 500, 2, 2588, 1)$$

Table 2. Value of Q_i^+ for each participant of Task 1

Participants	Q_i^+		Participants	Q_i^+
1	0.481154071		10	0.459536143
2	0.486380377		11	0.481035623
3	0.425769924		12	0.503450868
4	0.512654617		13	0.499094489
5	0.488922997		14	0.619961894
6	0.509868043		15	0.505601337
7	0.479417731		16	0.434555178
8	0.51068885		17	0.603533483
9	0.441772416		18	0.479186991

Table 3. The ranking of participants of Task 1

Order	Participants		Order	Participants
1	20		10	2
2	23		11	1
3	4		12	15
4	9		13	24
5	7		14	8
6	21		15	14
7	16		16	12
8	17		17	22
9	6		18	3

In this example, the estimation result is not matching with the actual selection.

In this experiment, Phase 3 was repeated on all the 348 tasks. Then the result of the estimation was compared with the actual selection made by the task provider to find out whether they are matchable. The latter is openly available on the crowdsourcing website.

4.3 Result Analysis

In order to evaluate the accuracy of the proposed method, the result of estimation is compared with the actual selection decision made by task providers. A metric of matching rate is used to indicate the percentages of matchable tasks. In this experiment,

the overall matching rate is 88.22 % (307 out of 348 tasks). The matching rates under different number of participants are presented in the following table.

The matching rates presented in Table 4 indicate that the proposed method has a high accuracy when the number of participants is between 4 and 10. There are 248 tasks (71.26 % of the total 348 tasks) fall into this range and the matching rates are above 90 %. Specially, in the simplest situation where only 4 participants competing a task, the estimation result is 100 % matched with the manual decision-making. This means the proposed method can achieve an estimation result that is very similar with the manual decision-making, when the manual decision-making is simple and the information overload problem does not appear.

Table 4. Matching rates under different number of participants

Number of Participants	Number of Tasks	Number of Matched tasks	Matching Rate	Number of Participants	Number of Tasks	Number of Matched tasks	Matching Rate
4	65	65	100%	31	11	8	72.73%
5	55	54	98.18%	32	11	8	72.73%
6	24	23	95.83%	37	8	5	62.5%
7	18	17	94.44%	38	3	3	100%
8	61	57	93.44%	39	5	3	60%
9	15	14	93.33%	40	4	2	50%
10	10	9	90%	42	2	1	50%
11	5	5	100%	44	2	1	50%
12	7	6	85.71%	47	2	1	50%
15	4	4	100%	52	3	1	33.33%
17	5	4	80%	57	3	1	33.33%
22	5	4	80%	60	1	0	0%
23	4	3	75%	62	3	1	33.33%
29	8	6	75%	63	4	1	25%

Technically speaking, the accuracy of this method will not be influenced by the number of participants. However, the matching rates generally declines along with the increase of participants. A possible explanation is that when the number of participants increase, it is more difficult to manually identify the best-fit participant. The task provider might have to rely on its own experience or even instinct in the decision-making for participant selection rather than an objective comparison between candidates. When the number of participants is more than 50, manually identifying the best-fit participant becomes very difficult and the actual selection result deviates very much from the estimation result, which results in matching rates below 33.33 %. The following curve demonstrate the decline of the matching rate. Under the assumption that the accuracy of

manual decision-making is mainly and negatively influenced by information overload, the proposed method could solve this problem and improve the decision-making for crowdsourcing tasks with a large number of participants. To solidly prove this statement, investigation on the factors that impact manual decision-making for selecting crowdsourcing participants is desired in the future (Fig. 1).

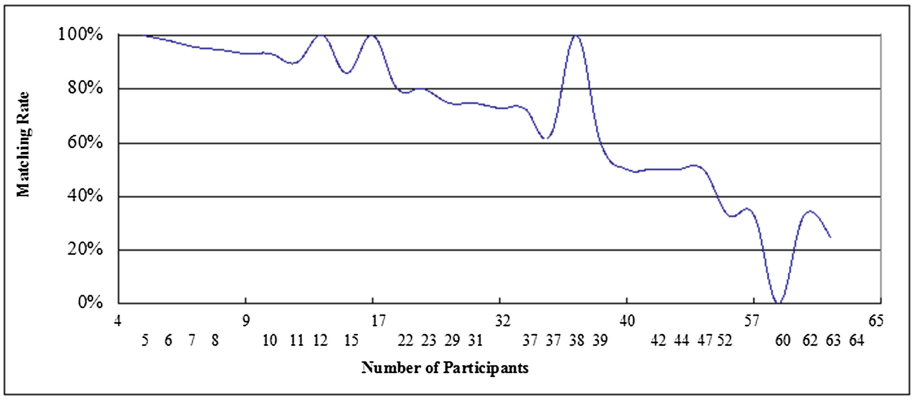


Fig. 1. The matching rates decline along with the increase of participants

5 Conclusions and Future Work

Companies want to adopt crowdsourcing to obtain on-demand IT services and enable flexible business processes. The challenge to overcome the information overload problem in estimating and identifying the best-fit crowdsourcing participant out of a large group of candidates. This paper presents a method of AHP-TOPSIS based on GRA for estimating crowdsourcing participants. The proposed method has been tested on a dataset of 348 valid and completed IT service crowdsourcing tasks from a Chinese crowdsourcing online marketplace platform. For the tasks with a small number of candidates, a matching between the estimation result and the actual selection made by task providers proofs the accuracy of the proposed method. Although the matching rate generally declines along with the increase of the number of candidates, this reflects the information overload problem in manual decision-making. The proposed method could solve this problem by providing an accurate and objective estimation which is uninfluenced by the number of participants. Employing this method allows task providers to quickly and easily identify the best-fit crowdsourcing participant. Furthermore, this method facilitates the design of a software agent for crowdsourcing platforms to rank the participants of a task automatically.

In the future work, it would be interesting to contact the task providers who had to make decision with a large number of candidates, especially those chose a participant other than the one recommended by the proposed method. Then invite them to interviews or surveys with questions such as how long it took them to make a decision on selecting the best-fit participant, what are the factors impacting their decision, and whether they think this method would help them in overcoming the information overload problem.

References

1. Howe, J.: The rise of crowdsourcing. Wired Mag. **14**(6), 1–4 (2006). Condé Nast
2. Lakhani, K.R.: Using the crowd as an innovation partner. Harvard Bus. Rev. **91**, 60–69 (2013)
3. Afuah, A., Tucci, C.L.: Crowdsourcing as a solution to distant search. Acad. Manage. Rev. **37**, 355–375 (2012)
4. Gong, Y., Janssen, M.: From policy implementation to business process management: principles for creating flexibility and agility. Gov. Inf. Q. **29**, 61–71 (2012)
5. Brabham, D.C.: Crowdsourcing as a model for problem solving. Convergence: Int. J. Res. New Media Technol. **14**, 75–90 (2008)
6. Prpic, J., Shukla, P.P., Kietzmann, J.H., McCarthy, I.P.: How to work a crowd: developing crowd capital through crowdsourcing. Bus. Horiz. **58**, 77–85 (2015)
7. Simula, H., Ahola, T.: A network perspective on idea and innovation crowdsourcing in industrial firms. Ind. Mark. Manage. **43**, 400–408 (2014)
8. Djelassi, S., Decoopman, I.: Customers' participation in product development through crowdsourcing: issues and implications. Ind. Mark. Manage. **42**, 683–692 (2013)
9. Gatautis, R., Vitkauskaite, E.: Crowdsourcing application in marketing activities. Procedia – Soc. Behav. Sci. **110**, 1243–1250 (2013)
10. Majchrzak, A., Malhotra, A.: Towards an information systems perspective and research agenda on crowdsourcing for innovation. J. Strateg. Inf. Syst. **22**, 257–268 (2013)
11. Satzger, B., Psaier, H., Schall, D., Dustdar, S.: Auction-based crowdsourcing supporting skill management. Inf. Syst. **38**, 547–560 (2013)
12. Nevo, D., Kotlarsky, J.: Primary vendor capabilities in a mediated outsourcing model: can IT service providers leverage crowdsourcing? Decis. Support Syst. **65**, 17–27 (2014)
13. Gefen, D., Gefen, G., Carmel, E.: How project description length and expected duration affect bidding and project success in crowdsourcing software development. J. Syst. Softw. (2015)
14. Tarasov, A., Delany, S.J., Namee, B.M.: Dynamic estimation of worker reliability in crowdsourcing for regression tasks: making it work. Expert Syst. Appl. **41**, 6190–6210 (2014)
15. Schall, D., Satzger, B., Psaier, H.: Crowdsourcing tasks to social networks in BPEL4People. World Wide Web **17**, 1–32 (2014)
16. Hwang, C.-L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications a State-of-the-Art Survey. Springer, Berlin (1985)
17. Saaty, T.L.: The Analytic Hierarchy Process: Planning, Priority Setting. Resource Allocation, Mcgraw-Hill (1980)
18. Ngai, E.W.T., Chan, E.W.C.: Evaluation of knowledge management tools using AHP. Expert Syst. Appl. **29**, 889–899 (2005)
19. Patil, S.K., Kant, R.: A fuzzy AHP-TOPSIS framework for ranking the solutions of knowledge management adoption in supply chain to overcome its barriers. Expert Syst. Appl. **41**, 679–693 (2014)
20. Deng, J.: Introduction to grey system. J. Grey Syst. **1**, 1–24 (1989)
21. Chen, M.-F.: Combining grey relation and TOPSIS concepts for selecting an expatriate host country. Math. Comput. Model. **40**, 1473–1490 (2004)

Mining Learning Processes from FLOSS Mailing Archives

Patrick Mukala[1(✉)], Antonio Cerone[1,2], and Franco Turini[1]

[1] Department of Computer Science, University of Pisa, Pisa, Italy
{mukala, cerone, turini}@di.unipi.it
[2] IMT Institute for Advanced Studies, Lucca, Italy
antonio.cerone@imtlucca.it

Abstract. Evidence suggests that Free/Libre Open Source Software (FLOSS) environments provide unlimited learning opportunities. Community members engage in a number of activities both during their interaction with their peers and while making use of these environments. As FLOSS repositories store data about participants' interaction and activities, we analyze participants' interaction and knowledge exchange in emails to trace learning activities that occur in distinct phases of the learning process. We make use of semantic search in SQL to retrieve data and build corresponding event logs which are then fed to a process mining tool in order to produce visual workflow nets. We view these nets as representative of the traces of learning activities in FLOSS as well as their relevant flow of occurrence. Additional statistical details are provided to contextualize and describe these models.

Keywords: FLOSS learning processes · Learning activities in open source · Mining software repositories · Process mining · Semantic search

1 Introduction

Currently a number of studies provide evidence that suggests the existence of learning opportunities in FLOSS environments [1, 10, 12–17, 22]. As part of this substantiation, FLOSS communities have been established as environments where successful collaborative and participatory learning between participants occurs [14, 16, 17].

Moreover, the levels of interest as well as the aura created around the occurrence of learning within FLOSS have attracted practitioners in tertiary education to consider incorporating participation in FLOSS projects as a requirement for some Software Engineering courses [12, 14, 24]. A number of pilot studies have been conducted in order to evaluate the effectiveness of such an approach in traditional settings of learning [10–13, 20]. To aid in this endeavor, in our previous study, we put an emphasis on how learning occurs in terms of phases [2, 19]. To this end, it has been proposed that a typical learning process in FLOSS occurs in three main phases: Initiation, Progression and Maturation. In each phase, a number of activities are executed though interactions between Novices and Experts. A Novice is considered as any participant in quest of knowledge while the knowledge provider is referred to as the Expert. Figure 1 depicts the categorization of the learning phases with the Initiation Phase synonymously

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 287–298, 2015.
DOI: 10.1007/978-3-319-25013-7_23

corresponding to understanding on the x axis as a learning stage, while Progression and Maturation correspond to practicing and developing respectively. The gray area in Fig. 1 represents the progression with regards to users as they progressively perform the types of activities on the y axis.

In this paper, we present an approach for mining these learning phases from FLOSS data. For illustrative purposes, we detail our approach and present the results for the understanding (Initiation) phase, which is at the bottom of the scale in Fig. 1. In this phase, FLOSS participants get involved in the projects by reviewing and communicating with the purpose of understanding contents without producing any tangible contributions. Initiation is a critical stage as the participant accesses project repositories and exchanges emails and posts messages seeking information and posting any requests. Figure 1 also shows how, in the practicing and developing phases, the participants' activities gradually move from simply using to posting and making significant contributions through commits [2].

FLOSS repositories, such as CVS, Bug reports, mailing archives, Internet relay chats etc., contain all traces of participants' activities as they work in these environments. Singh et al. [6] argue that the FLOSS environment typically includes discussion forums or mailing lists to which users can post questions and get help from developers or other users. These forums are unrestricted and act as a learning environment for novices and experts alike. While many studies have provided invaluable insights in this direction, their results are mostly based on surveys and observation reports [7, 8, 21–23, 24]. Our paper proposes to contribute in this context by studying learning activities from FLOSS repositories using process mining. In particular, the paper focuses on tracing and visualizing the learning activities as well as their flow of occurrence collected in mailing archives of a FLOSS platform called OpenStack [25].

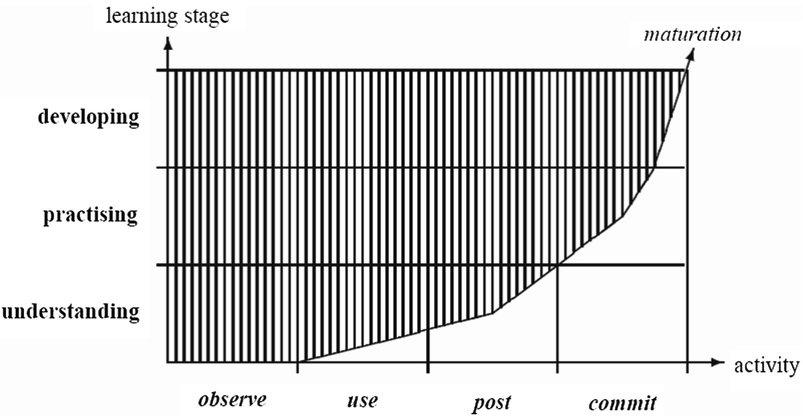


Fig. 1. Learning stages and participants' learning progression in OSS communities [2]

Our major contribution is the approach used in analyzing the data, the application of process mining and mostly the discovered empirical evidence of learning activities' traces. The rest of the paper is structured as follows. Section 2 provides preliminary

details on mining the data and constructing the log and succinctly describes the Initiation Phase of the learning process. In Sect. 3 we discuss the data collection and analysis and then present the empirical results. Section 4 concludes the paper.

2 Preliminaries: Mining Data and Catalog of Key Phrases

In order to identify activities and construct the event logs needed for our analysis, we undertake a number of tasks. The first task is analyzing the contents of emails. Text mining appears to be the most direct solution for this task as we need to analyze the contents of a post/email and deduct a corresponding activity. Current text mining tools such as Carrot2, GATE, OpenLP, RapidMiner and KH Coder appear not to be appropriate for the kind of analysis we want to conduct. Tracing learning activities requires semantic interpretation of email contents and this could not be achieved by using any of these tools.

Therefore, we considered making use of Semantic search with MS SQL as a fit alternative. Semantic search improves search by understanding the contextual meaning of the terms and tries to provide the most accurate answer for a given text document. However, this also requires the use of key phrases to steer the search [18]. Our choice of key phrases is based on a number of studies conducted in FLOSS with regards to the kinds of questions and answers that are asked in FLOSS communication environments [3–5, 27]. We start from this categorization, following questions and responses categories; then we deduct a number of key phrases. We try as much as we can to include all the identified key phrases and expressions within the context of identifying learning activities and establishing the learning process across its three phases, although in this paper we only present the details of the first phase.

We make use of previous findings [3, 4], a formal model of learning activities in FLOSS communities [19], as well as lexical semantics to draw a catalog of key phrases with respect to our endeavor. Lexical semantics builds from synonyms of terms and their homonyms to derive the meaning of words in specific contexts. Hence, making use of semantic search is paramount and promises to capture the meaning of message contents as much as possible in identifying activities. Figure 2 presents a catalog that contains the key phrases that semantically identify activities as categorized according to the participants' roles in the Initiation Phase of the learning process.

Principal activities gravitate around observing and making contacts in the Initiation Phase of the learning process [19]. Ideally, this step constitutes an opportunity for the Novice to ask questions and get some help depending on the requests while the Expert intervenes at this point to respond to such requests.

On the one hand, a Novice seeking help can execute a number of activities. These include *FormulateQuestion*, *IdentifyExpert*, *PostQuestion*, *CommentPost* or *PostMessage*, *ContactExpert* and *SendDetailedRequest*. On the other hand, the main activities as undertaken by the Expert during the same period of time include *ReadMessages* on the mailing lists/Chat messages, *ReadPost* from forums, *ReadSourceCode,* as any participant commits code to the project, or *CommentPost*, *ContactNovice* and *CommentPost*.

STATES	GLOBAL KEYWORDS	PARTICIPANTS	ACTIVITIES	KEYPHRASE/CONDITIONAL ACTIVITY
Observation	"problem", "help", "error"	NOVICE	FormulateQuestion	If PostQuestion = true
			IdentifyExpert	"How did you do this", "I saw your code", "I need your help", "this does not work for me", "-1", "is this possible to do this", "can this be done?", "very helpful", "very well"
			PostMessage	If IdentifyExpert = true
			PostQuestion	"How can I do…?" "How to?", "don't understand how", "could help?", "what is wrong?", "my code is not running", "code not executing", "question", "How to", "what is wrong", "where can I", "Any ideas how to solve this problem?", "I have tried doing", "search for this", "but have had little luck", "any help?", "any suggestions?", "everything I could", "new to the"
			CommentPost	"does not work", "not executing", "this does not work for me", "do not know what is wrong with my code", "here is my code", "in short my problem", "step by step", "details provided", "works as follows", "I want it to", "expect it to", "my question is like this", "what I mean"
		EXPERT	ReadMessages	If CommentPost = true
			ReadPost	If CommentPost = true
			ReadSourceCode	"syntax error", "maybe you should…", "it seems to work for me""do not know what is wrong with the code" or "not sure it can"") or "running your code"
			CommentPost	"system details needed", "more details needed", "more problems details needed", "more details needed of what is on the screen", "Did it work before?", "provide exact step by step details"
ContactEstablishement	"can I get your help", "can you help", "send question", "contact details", "send email", "send file", "more details"	NOVICE	ContactExpert	If SendDetailedRequest = true
			SendDetailedRequest	"actually the code is like this…" "I tried this", "I don't know how", "I don't understand how?" "can you help", "your help", "you explain", "as you asked", "so my question is", "I wanted to know", "what I meant is", "my screenshot looks", "I get this error", "how do I fix this"
		EXPERT	ContactNovice	If CommentPost/SendFeedback = true
			CommentPost/SendFeedback	"does not work", "not executing", "maybe you should…", "it seems to work for me", "this does not look right", "you need to delete this.", "the syntax is not correct", "send me your code", "what is your problem?", "this works for me", "Did it work before", "I think it should work"

Fig. 2. Catalog of key phrases for initiation phase

In order to conduct our analysis, we need to identify the most appropriate repository in this regard. The main criteria in making such a decision lies on the existence of some form of communication exchange between FLOSS members on any candidate repository. Mailing Archives contain email messages between FLOSS members about discussions on topics relevant to the community. Some of these topics involve general questions or specific requests about files, pieces of code or even the use of new plug-ins etc. Hence, these Mailing Archives provide adequate details to track activities and explain their flow of occurrence in the Initiation Phase. Moreover, it is worth noting that the same approach can be applied to mine the remaining phases on other repositories such as source code or commits.

3 Data Collection and Analysis

The FLOSS platform used in our analysis is OpenStack [25]. According to Wikipedia, "OpenStack is a free and open-source software cloud computing software platform. Users primarily deploy it as an infrastructure as a service (IaaS) solution. The technology consists of a series of interrelated projects that control pools of processing, storage, and networking resources throughout a data center—which users manage through a web-based dashboard, command-line tools, or a RESTful API that is released under the terms of the Apache License" [25].

We considered this platform mainly due to the availability of data about email archives and also because it is still an active platform. This database is made up of 7 tables that store data pertaining to compressed files (source_code file, bugs), the mailing lists as per group discussions and topic of interests, the number of messages exchanged as well as details of the individuals involved in these exchanges as shown in Table 1.

This repository contains exactly 54762 emails exchanged between 3117 people who are registered on 15 distinct mailing lists. These emails were sent during a period of time spanning from 2010 to 2014. The length of the messages considered is of typical email length specifically with an average of 3261 characters, the longest email was of 65535 characters and the shortest message yields a single character length.

In order to analyze this data set, we make use of process mining techniques. The key in Process Mining is to identify events. An event is a tuple made up essentially of case ID, performer, activity and any relevant attributes we need for our analysis. In our case, we include the phase of the learning process, date as well as the role (Novice, Expert). Other key components include the catalog of key phrases as shown in Fig. 2 and the data set. Based on all these elements, we generated our event log, which is the set of all identified events.

An event E is a sextuple (t, a, p, d, s, r) such that: t is the case in the event and can be either a topic on emails or an issue number on code and bug reports; a is the activity; p is the participant; d is the relevant date of occurrence; s is the state of the learning process; and r is the participant's role in the process.

Moreover, we refer to the catalog introduced earlier to retrieve the mappings between key phrases, activities, states and participants. Let c_1, c_2 and c_3 be catalogs respectively for Initiation, Progression and Maturation. We distinguish between key

Table 1. Details of mailing archives elements from OpenStack

Tables	Records
dbo.compressed_files	401
dbo.mailing_lists	15
dbo.mailing_lists_people	4434
dbo.messages	54762
dbo.messages_people	54762
dbo.people	3117
dbo.people_upeople	3117

phrases for activities and states. We refer to key phrases for states as *gl_key* (global keys) while the key phrases that help distinguish activities are referred to as *lc_key* (local keys). We define catalogs as sextuples (C, c_i, *gl_key, state, lc_key, activity, role*) such that: C is the set of all our catalogs, $c_i \in C$ is a single catalog, *gl_key* is the key phrase for the identification of a state, *state* is the state as it appears in the catalog, *lc_key* is the key phrase used to identify an activity, *activity* is the corresponding activity in the catalog, and *role* is the role as it appears in the catalog. Using such information, we generate the event log to be analysed through process mining.

3.1 Process Mining Mailing Archives

In order to process mine these records, we choose Disco (Discover Your Processes) [26], an appropriate tool for analyzing the identified events and providing efficient visualizations to demonstrate the workflow of occurrence of activities in these processes. Disco is a toolkit for process mining that enables the user to provide a preprocessed log specifying case, activities, originator and any other needed attributes. The tool performs automatic process discovery from the log and outputs process models (maps) as well as relevant statistical data.

In essence, Disco applies process mining techniques in order to construct process models based on available logging data that is organized into an event log. This logging data is all the details about transactions that can be found in log file or transaction databases. Therefore, an event log can take a tabular structure containing all recorded events that relate to executed business activities [26].

Making use of Disco, we produced the Process Models representing the occurrence of learning activities as documented by their corresponding email messages.

For simplicity, we choose to represent the models through a graphical workflow as well as the statistical information as provided by Disco. Disco offers the possibility for a process model to be represented with frequency metrics that explain the flow of occurrence of events.

The main objective of the frequency metrics is the depiction of how often certain parts of the processes have been executed. We can distinguish three levels of frequency: absolute frequency, case frequency and maximum repetitions. We consider these metrics to model learning activities executed by both the Novices and Expert.

Additional details regarding the numerical measures such as events over time, active cases during this given period of time, case variants, the number of events per case as well as case duration could be plotted as needed. However, for simplicity and effectiveness, we represent only major statistical details that are most representative of the presence, impact and occurrence of learning activities in FLOSS over the chosen period of time.

3.2 Empirical Results

Before we unpack details about process models for both the Novice and Experts, we give some crucial details about the overall Initiation Phase. It should be noted that in Figs. 3 and 4 the numbers, the thickness of the arcs or edges, and the coloring in the model illustrate how frequent each activity or path has been performed. For the purpose of this paper, we retained the topic of emails, the message itself, the people involved in exchanging these emails, the resulting activities and classification of where such activities fall in our defined learning curve to build events and produce the event log used for model extraction.

The analysis of the Initiation Phase of the learning process is carried out on data that refer to the period between the 11th of November 2010 and the 6th of May 2014. During this time, we note that a total of 123401 events were generated. An event represents a tuple made up of the case (in this context, the discussion topic), the email senders as well as the relevant learning activities. With about 565 cases, a total of 14 activities are executed with an average time per case of 69.9 days while the median duration is of 57.8 days.

We can also point out that participants in quest for knowledge claim the majority of activities with a total of 122838 amounting to 99.54 % of all executed activities at this point in contrast with Experts who intervene at a lower rate of 0.36 % with 440 activities, slightly ahead of people doing something other than exchanging knowledge with 123 activities.

The process model depicted in Fig. 3 represents a workflow for all the activities performed by the Novice during the first phase of the learning process. On average, how often an activity has been executed in this process by the Novice as well as how often an activity links to another (path) can be noted through the numbers, the thickness of the arcs or edges, and the coloring in the model. We note that the Novice in OpenStack has engaged in a number of learning activities throughout this period of time. Figure 3 demonstrates that in 51 cases the process would start from formulating a question, posting the question, commenting on post (and this has occurred about 27 times), posting a message, which indicates that an expert has been identified, contacting that expert and sending a detailed request to the expert through commenting on a post. The numerical argument between the transitions from one activity to another indicates how many times on average this has happened.

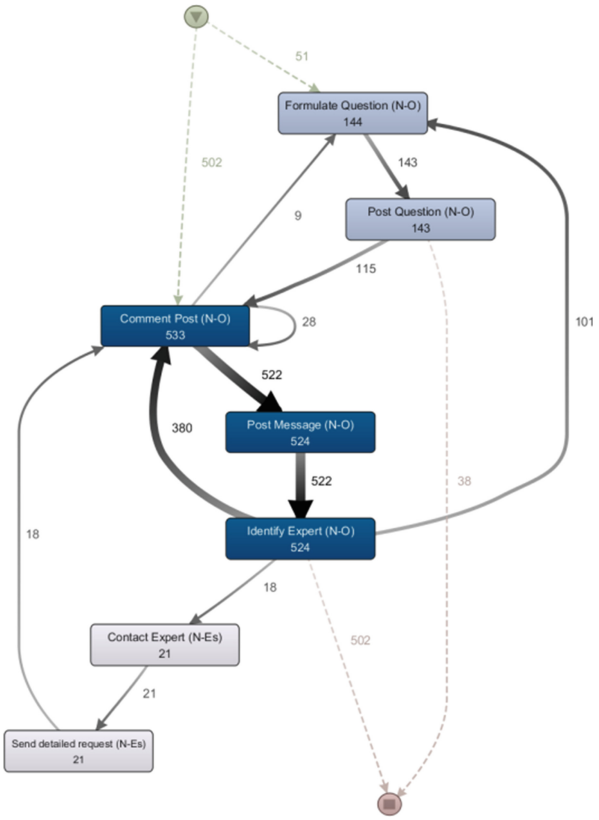


Fig. 3. Process model for novice–per frequency [initiation phase] in mailing lists

Moreover, Fig. 3 also shows that in 502 cases, a Novice starts by commenting on a post first, then formulates a question and follows the process as explained above. Sometimes (101 times in the depicted process map), after identifying an expert, a Novice could go back to formulating another question (or follow-up questions) or even go back to just commenting on the post as part of the interaction.

The process model depicted in Fig. 4 represents a workflow for all the activities performed by the Expert during the first phase of the learning process. One should note that 6 main activities are undertaken by the Expert. In some cases, the Expert would contact the Novice, by commenting on a post or giving feedback regarding a request from the Novice, then read messages and posts, commenting on these posts as well as reading source code, especially if the Novice's requests have to do with source code. The assumption here is that the Expert is referred to as such because of the nature of the reaction activity. Every time an Expert comments, it is in response to a Novice request or to request further details on an already posted question.

In some instances, the Expert would go back to reading messages after commenting on a post or reading source code, or sometimes contacting the Novice again after commenting on a post involved in the exchange. In some instances, on average in 9

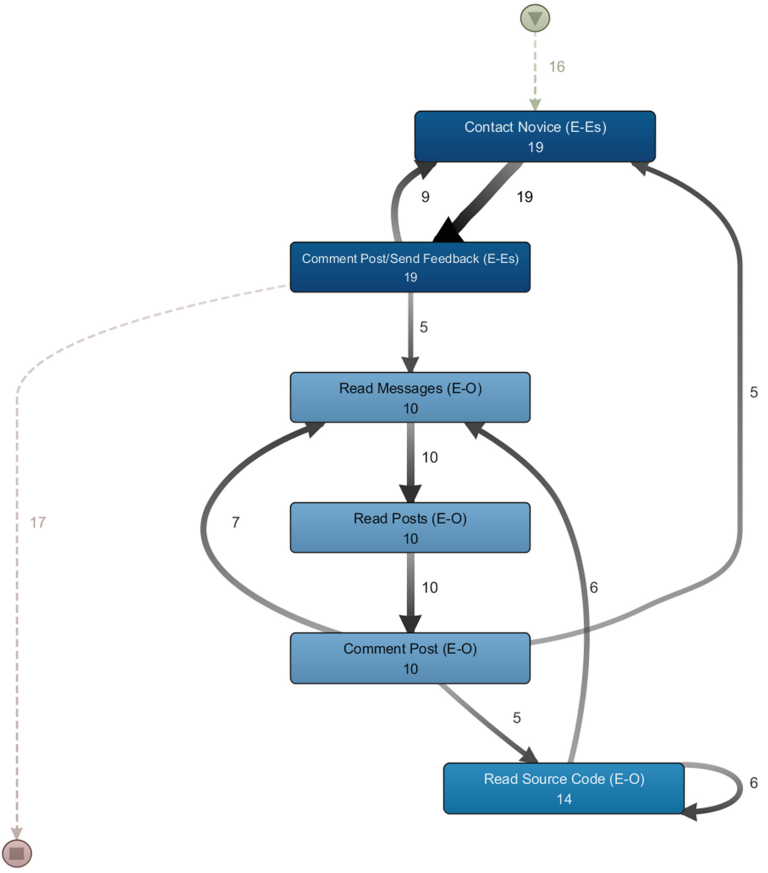


Fig. 4. Process model for expert–per frequency [initiation phase] in mailing lists

cases, the Expert goes back to contacting the Novice after providing feedback. This is where further details about the request or clarification might be requested.

4 Conclusion

FLOSS environments appear to provide learning opportunities for participants. While this aspect has been previously investigated [7, 8], we believe that there is a need for empirical support for such findings. An important remark [9] emphasizes that in such previous work, content data had been collected using surveys and questionnaires or through reports from observers who have been part of the community for a defined period of time. Research suggests that a growing number of participants are highly motivated to engage in these platforms through discussions and email exchange. These interactions produce massive volumes of data that contain evidence of learning. Since these learning activities are not directly observable in the repositories, we made use of semantic search in MS SQL to identify activities from message texts and constructed the event log that we then analyzed with the help of the Disco process mining tool.

Using a combination of text mining techniques (semantic search), key phrases and a set of rules, we believe that our approach has provided insights on how to find traces of interaction and learning activities in FLOSS email messages. We can thus say that it is feasible to trace these activities and that process mining can play a catalyst role in identifying these activities in FLOSS. Our work aimed to give evidence about the existence of learning processes in FLOSS environments through the empirical analysis of OpenStack Mailing Archives.

Our results demonstrate how these learning processes are extracted and how each activity fits within the global picture. For a Novice, we can note that the process in some cases spans from formulating a question, posting the question, commenting on post, posting a message, which indicates that an expert has been identified, contacting that expert and sending a detailed request to the expert through commenting on a post. In most cases, a Novice starts by commenting on a post first, and then formulates a question before following the process as described above. Six main activities are performed by the Expert starting from contacting the Novice, through commenting on a post or giving feedback regarding a request from the Novice. The Expert will then read messages and posts, commenting on these posts as well as reading source code, especially if the Novice's requests have to do with source code. In some instances, the Expert would go back to reading messages after commenting on a post or reading source code, or sometimes contacting the Novice again after commenting on a post involved in the exchange. In some other instances the Expert goes back to contacting the Novice after providing feedback. This is where further details about the request or clarification might be requested.

Finally, more experiments using this approach can be conducted in order to trace activities in the next two phases. These phases include the Progression and Maturation phases. Figure 1 indicates how FLOSS contributors start by just observing, in the Initiation Phase, and gradually evolve to more tangible activities such as posting and committing software artifacts and source code [2]. The types of activities in the Progression phase include *Revert*, *Post* and *Apply*. After the Expert makes contact, the Novice provides additional details about the previous request through activity *Revert.* The Expert gets further involved by providing guidance and the required help to the Novice through *Post,* while the Novice implements the new acquired knowledge through *Apply*. During the Maturation phase, activities include *Analyze, Commit, Develop, Review* and *Revert*. In this phase, the Novice's acquired skills are expressed through the execution of advanced activities such as producing new code, reviewing new commits and providing assistance during discussions, thus gradually performing a transition from Novice to Expert [19]. The Expert performs the same activities with an emphasis on transferring knowledge and providing help when needed rather than applying new skills.

References

1. Sowe, S.K., Stamelos, I.: Reflection on knowledge sharing in F/OSS projects. In: Russo, B., Damiani, E., Hissam, S., Lundell, B., Succi, G. (eds.) Open Source Development, Communities and Quality. IFIP, vol. 275, pp. 351–358. Springer, Heidelberg (2008)

2. Cerone, A.: Learning and activity patterns in OSS communities and their impact on software quality. In: Proceedings of the 5th International Workshop on Foundations and Techniques for Open Source Software Certification, (OpenCert 2011). Electronic Communications of the EASST, vol. 48. The European Association of Software Science and Technology (2011)
3. Lakhani, K.R., Von Hippel, E.: How open source software works: "free" user-to-user assistance. Res. Policy **32**(6), 923–943 (2003)
4. Steehouder, M.F.: Beyond technical documentation: users helping each other. In: Proceedings of Professional Communication Conference, IPCC 2002, pp. 489–499. IEEE International (2002)
5. Singh, V., Twidale, M.B., Nichols, D.M.: Users of open source software-how do they get help? In: hicss, pp. 1–10). IEEE, December1899
6. Singh, V., Twidale, M.B., Rathi, D.: Open source technical support: a look at peer help-giving. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences, HICSS 2006, vol. 6, pp. 118c–118c. IEEE, January 2006
7. Glott, R., Meiszner, A., Sowe, S.K., Conolly, T., Healy, A., Ghosh, R., Karoulis, A., Magalhães, H., Stamelos, I., Weller, M.J., West, D.: FLOSSCom-Using the Principles of Informal Learning Environments of FLOSS Communities to Improve ICT Supported Formal Education
8. Glott, R., Meiszner, A., Sowe, S.K.: FLOSSCom phase 1 report: analysis of the informal learning environment of FLOSS communities. FLOSSCom Project (2007)
9. Cerone, A., Fong, S., Shaikh, S.A.: Analysis of collaboration effectiveness and individuals' contribution in FLOSS communities. In: Proceedings of the 5th International Workshop on Foundations and Techniques for Open Source Software Certification (OpenCert 2011). Electronic Communications of the EASST, vol. 48. The European Association of Software Science and Technology (2011)
10. Papadopoulos, P.M., Stamelos, I.G., Meiszner, A.: Enhancing software engineering education through open source projects: four years of students' perspectives. Educ. Inf. Technol. **18**(2), 381–397 (2013)
11. LeBlanc, R.J., Sobel, A., Diaz-Herrera, J.L., Hilburn, T.B.: Software Engineering 2004: Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering. IEEE Computer Society Press, Washington, D.C. (2006)
12. Sowe, S.K., Stamelos, I.G.: Involving software engineering students in open source software projects: experiences from a pilot study. J. Inf. Syst. Educ. **18**(4), 425 (2007)
13. Cerone, A.K., Sowe, S.K.: Using free/libre open source software projects as e-learning tools. In: Proceedings of the 4th International Workshop on Foundations and Techniques for Open Source Software Certification (OpenCert 2010). Electronic Communications of the EASST, vol. 33. The European Association of Software Science and Technology (2010)
14. Fernandes, S., Cerone, A., Barbosa, L.S.: Analysis of FLOSS communities as learning contexts. In: Counsell, S., Núñez, M. (eds.) SEFM 2013. LNCS, vol. 8368, pp. 405–416. Springer, Heidelberg (2014)
15. Fernandes, S., Cerone, A., Barbosa, L.S.: FLOSS Communities as Learning Networks. Int. J. Inf. Educ. Technol. **3**(2), 278–281 (2013)
16. Fernandes, S., Cerone, A., Barbosa, L.S., Papadopoulos, P.M.: FLOSS in technology-enhanced learning. In: Cerone, A., Persico, D., Fernandes, S., Garcia-Perez, A., Katsaros, P., Ahmed Shaikh, S., Stamelos, I. (eds.) SEFM 2012 Satellite Events. LNCS, vol. 7991, pp. 121–132. Springer, Heidelberg (2014)
17. Meiszner, A., Glott, R., Sowe, S.K.: Free/libre open source software (FLOSS) communities as an example of successful open participatory learning ecosystems. UPGRADE Eur. J. Inform. Prof. **9**(3), 62–68 (2008)

18. Larson, B.: Delivering Business Intelligence with Microsoft SQL Server 2012. McGraw-Hill Osborne Media, New York (2012)
19. Mukala, P., Cerone, A., Turini, F.: An abstract state machine (ASM) representation of learning process in FLOSS communities. In: Canal, C., Idani, A. (eds.) SEFM 2014 Workshops. LNCS, vol. 8938, pp. 227–242. Springer, Heidelberg (2015)
20. Meiszner, A., Mostaka, K., Syamelos, I.: A hybrid approach to computer science education–a case study: software engineering at Aristotle University. In: CSEDU 2009 - Proceedings of the First International Conference on Computer Supported Education, Lisboa, Portugal (2009)
21. Dillon, T., Bacon, S.: The potential of open source approaches for education. FutureLab opening education reports (2006). http://www.futurelab.org.uk/resources/publicationsreports-articles/opening-education-reports/Opening-Education-Report200
22. Sowe, S.K., Stamelos, I., Deligiannis, I.: A framework for teaching software testing using F/OSS methodology. In: Damiani, E., Fitzgerald, B., Scacchi, W., Scotto, M., Succi, G. (eds.) Open Source Systems. IFIP, vol. 203, pp. 261–266. Springer, Heidelberg (2006)
23. Sowe, S.K., Stamelos, I., Angelis, L.: An empirical approach to evaluate students participation in free/open source software projects. In: IADIS International Conference on Cognition and Exploratory Learning in Digital Age CELDA 2006, pp. 304–308 (2006)
24. Meiszner, A., Stamelos, I., Sowe, S.K.: 1st international workshop on: 'designing for participatory learning' building from open source success to develop free ways to share and learn. In: Boldyreff, C., Crowston, K., Lundell, B., Wasserman, A.I. (eds.) OSS 2009. IFIP AICT, vol. 299, pp. 355–356. Springer, Heidelberg (2009)
25. OpenStack: In Wikipedia, the free encyclopedia, 3 November 2014. http://en.wikipedia.org/w/index.php?title=OpenStack&oldid=632224644. Accessed 10 November 2014
26. Günther, C.W., Rozinat, A.: Disco: discover your processes. In: Proceedings of the Demonstation Track of BPM 2012. CEUR Workshop Proceedings, vol. 940, September 2012. CEUR-WS.org
27. Elliott, M.S., Scacchi, W.: Free software development: cooperation and conflict in. Free/open Source Softw. Dev. **152** (2005)

Private-Collective Innovation and Open Source Software: Longitudinal Insights from Linux Kernel Development

Dirk Homscheid[1,2(✉)], Jérôme Kunegis[2], and Mario Schaarschmidt[1,2]

[1] Institute for Management, University of Koblenz-Landau, Koblenz, Germany
[2] Institute for Web Science and Technologies,
University of Koblenz-Landau, Koblenz, Germany
{dhomscheid, kunegis,
mario.schaarschmidt}@uni-koblenz.de

Abstract. While in early years, software technology companies such as IBM and Novell invested time and resources in open source software (OSS) development, today even user firms (e.g., Samsung) invest in OSS development. Thus, today's professional OSS projects receive contributions from hobbyists, universities, research centers, as well as software vendors and user firms. Theorists have referred to this kind of combined public and private investments in innovation creation as private-collective innovation. In particular, the private-collective innovation model seeks to explain why firms privately invest resources to create artifacts that share the characteristics of non-rivalry and non-excludability. The aim of this research is to investigate how different contributor groups associated with public and increasing private interests interact in an OSS development project. The results of the study show that the balance between private and collective contributors in the Linux kernel development seems to be changing to an open source project that is mostly developed jointly by private companies.

Keywords: Open source · Open source community · Private-Collective innovation · Linux kernel

1 Introduction

Open source software (OSS) has changed how researchers and practitioners look at software development and related business models [1–3]. OSS differs from traditional in-house software development, as its outcome, the software, is freely accessible by anyone. As a result, major companies such as Facebook, Google and Twitter use open source technologies. In 2012 one million open source projects were catalogued and this figure was projected to double within two years [4]. In addition, according to IDC research, the OSS market should be worth approx. $8 billion in 2013 [5].

Due to this success, firms have also started to actively engage in OSS development [6, 7]. While in early years, software technology companies such as IBM and Novell invested time and resources in OSS development, today even user firms (e.g., Samsung)

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 299–313, 2015.
DOI: 10.1007/978-3-319-25013-7_24

invest in OSS development [8]. Thus, today's successful OSS projects receive contributions from hobbyists, universities, research centers, as well as from software vendors and user firms [9, 10]. Theorists have referred to this kind of combined public and private investments in innovation creation as private-collective innovation [11]. This concept asserts that a private investment model – where firms create and commercialize ideas themselves – and a collective invention model, where multiple economic actors create public goods innovations, may coexist under certain circumstances [12]. In particular, the private-collective innovation model seeks to explain why firms privately invest resources to create artifacts that share the characteristics of non-rivalry and non-excludability [13].

The private-collective model also implicitly assumes that private and public investments in innovations are approximately equal. However, successful OSS projects receive more than 75 % of their code from contributors[1] who are paid by a company [8] and the majority of code is written between 9 am and 5 pm – again indicating that contributions are predominantly provided by firms [14]. These figures contrast with the picture of private-collective innovation as an invention mode where public and private interests manifest equally. The aim of this research therefore is to investigate how different contributor groups associated with public and increasing private interests interact in an OSS development project.

In order to study the interplay of both interest groups we not only need to consider demographic characteristics of the community but also the structural patterns of interactions in it. To achieve this goal, we analyze developers active in the Linux kernel (LK) development community from a social network point of view, as the interaction between the members of a software development community reflects the structure of their collaboration. In particular, we investigate degree distributions and the Gini coefficient in the contributor network with respect to the private and collective contributor groups. Network centrality measures are important indicators of influence in OSS development and are known to deviate according to having firm sponsorship or not [15].

We start with detailing what volunteer and firm-sponsored (i.e., employed) developers motivate to participate in OSS development. Then, we discuss the private-collective innovation model in more detail. Based on a dataset of mailing list communication of LK developers from 1996 to 2014 we calculate network measures for each type of developer (e.g., firm-sponsored, hobbyist, university-affiliated, etc.) and compare them for each year. We discuss implications for research and provide further avenues for research concerning private-collective innovation.

2 Theoretical Background

2.1 Open Source Software Contributors

OSS Communities. An OSS project relies on contributors who make up the core element of an OSS community. OSS is commonly understood as a type of software that

[1] In this study the terms *contributor*, *developer*, *actor*, *participant* and *programmer* are used synonymously to denote people who are active in OSS projects and in Linux kernel development.

can be used, changed, and shared by any person. The software itself is in most cases developed by a heterogeneous group of people and distributed under specific licenses, which guarantee the above-mentioned characteristics of OSS [16].

In general, a community arises when different people come together and share a common interest [17]. Thus, von Hippel and von Krogh [11] conceptualize OSS development communities as "*Internet-based communities of software developers who voluntarily collaborate in order to develop software that they or their organizations need*" [11, p. 209]. Besides the fact that OSS communities consist of hobbyists, who voluntarily provide their resources to the community, the definition also involves another important contributor group – organizations. Organizations differ from hobbyists in terms of their motivation to engage and are represented in the community by their employed developers. In turn, employed developers might be considered as proxies for firm interests in the community.

Motivation of Voluntary OSS Developers. The pertinent literature specifies intrinsic and extrinsic motivation as major drivers for hobbyists to engage in OSS projects (e.g., [18–21]). Intrinsic motivation is the execution of an activity due to the accompanying enthusiasm and not for the achievement of specific results [22]. A behavior is extrinsically motivated when an activity is performed for reward, recognition or because of an instruction from someone or an obligation [22]. Although researchers agree on different forms of intrinsic and extrinsic motivation, there is often disagreement about their relevance. The most relevant forms of both motivation types in the context of OSS developers are described briefly in the following.

In connection with OSS developers, researchers investigated a plethora of intrinsic motivators. Among these, *joy-based intrinsic motivation* is the strongest and most prevalent driver of OSS contributors [19]. Joy-based motivation is closely linked to the creativity of a person. Frequently, contributors to OSS projects have a strong interest in software development and related challenges [18].

Another fundamental aspect of intrinsic motivation is *altruism*, which is the desire to help others and to improve their welfare. In OSS communities, developers code programs, report bugs, etc., at their own expense, which includes the invested time and opportunity costs. They participate in the OSS community, without taking advantages of its outcome [18, 21, 23].

In addition, the *OSS ideology* plays a crucial role for many contributors and involves

- joint collaborative values, such as helping, sharing and collaboration,
- individual values, such as learning, technical knowledge and reputation,
- OSS process beliefs, such as code quality and bug fixing and
- beliefs regarding the importance of freedom in OSS, such as an open source code and its free availability and use for everyone [24].

Besides these distinguishing aspects of participants' intrinsic motivation to engage in OSS projects, researchers have found that extrinsic stimuli can also have an impact on the activities of actors in communities (e.g., [18–21]).

An extrinsic stimulus is given through a *personal need* of a developer. Many OSS projects are launched because the initiators needed software with specific functions that are not available to date, and they have the willingness and knowledge to develop these [21].

OSS communities offer the possibility for developers to *improve their programming skills* and their knowledge through participation in a project. Programmers are free to choose in which tasks they participate according to their interests and abilities. As a result, the self-learning participants are experiencing a continuous learning curve and build a repertoire of experiences, ways and means to solve specific software development tasks [21].

In addition, "signaling incentives" as described by Lerner and Tirole [25] can also be a reason for people to participate in OSS communities. The incentives cover, inter alia, the *recognition by other members* of the community and the *improvement of the professional status*.

Motivation of Firms Involved in OSS Development. In addition to hobbyists, companies are also active in OSS communities. While voluntary OSS contributors are driven by intrinsic and extrinsic values, economic and technological aspects motivate firms to participate in OSS projects. In recent times, companies open outwards to organize their innovation activities more effectively and efficiently. A means to complement their own resource base are innovation communities. In the case of software companies, OSS communities form a resource pool these firms can benefit from – depending on the strategy they pursue [6, 26]. Literature investigating motivational aspects of companies active in OSS projects reveals that economic theory is not sufficient to explain the relation between firms and their OSS engagement. Andersen-Gott et al. [27] have reviewed this issue and identified the following three categories of motivational factors that are relevant for companies active in OSS communities.

1. *Innovative Capabilities.* If the involvement of a company in an OSS project is aligned with the business model it maintains, the interaction with the community can lead to better or new products which imply a competitive advantage. The inclusion of external contributors increases the firm's innovative capacity.
2. *Complementary Services.* The dominant way for firms to appropriate from OSS is by providing complementary services to customers (e.g., training, technical support, consultancy and certifications [28]) aligned with their business strategy [29]. Firms pursuing this concept deploy own employees that also contribute to the open source project and community work. Thus, the company (1) acquires external knowledge through their own employees active in OSS development and (2) has access to complementary resources in the community, which are difficult to replicate internally [30].
3. *Cost Reduction.* Companies can publish the source code of their proprietary software under an OS license, try to attract external developers and build a community around the software. In this case, the company will get, for example, ideas for new features, bug reports, documentation, and extensions of the software from external contributors without having to pay for it [25, 31]. Further, in the long run, the code is maintained by the community, such that the firm has lower costs than its competitors with proprietary software [32]. However, it should be noted that establishing an ecosystem and an active community around released source code is no easy task as rivals could pursue similar strategies [26, 33].

2.2 Private-Collective Innovation

In organization science, two different modes of innovation are dominant, namely the private investment and the collective action model.

The *private investment model* is associated with a rather closed innovation behavior. Innovators tend to protect their internally developed proprietary knowledge as this is the source of their profits and competitive advantage [11]. Here innovation is clearly seen as a closed process driven by private investments in order to lead to private returns for the innovator [34].

The *collective action model* of innovation is connected to the provision of a public good. Innovators collaborate in order to develop a public good under conditions of market failure. The produced good is characterized by non-excludability and non-rivalry [35]. This model requires that innovators supply their collected knowledge about a project to a common knowledge base and thus make it a public good. This innovation method can unfortunately be exploited by free riders, who wait until other contributors have done the work and use the outcome for free [11, 35].

OSS communities are an example for a mixture of both mentioned innovation models. OSS contributors freely reveal their privately developed source code as a public good. The developers do not make commercial use of their property rights, although the source code is created as a result of private investments. This innovation behavior is termed private-collective innovation [11]. To get a deeper understanding of how OSS communities combine the best of both models, OSS innovation is first considered from the private investment and second from the collective action point of view.

From the *private investment model perspective*, OSS deviates in two major aspects from the conventional private investment model. First, software contributors are the actual innovators in OSS rather than commercial software developers, because they create software that is needed either by themselves or by the community. Second, OSS developers freely reveal the source code, which they have developed by private means; this manner stands in contrast to the classical innovation behavior. Due to the lack of a commercial market for the sale or licensing of OSS, it is made openly available as a public good [11]. Rewards for the developers are provided in forms other than money or commercialization of property rights. Contributors gain private profits such as reputation, experience or reciprocity [25, 36].

From the *collective action model view*, the community produces a public good with its attributes of non-excludability and non-rivalry [35]. Taking the above given description of the collective action model into account, the non-excludability would bring a dilemma with it because free riders benefit from the software but do not contribute to the good compared to the developers. This circumstance is not a problem, as in line with the OS ideology people voluntarily participate in OSS development and share the results without costs [11]. Moreover, contributors obtain benefits, for example problem solving expertise, learning and enjoyment, from the participation on developing a public good, which the free rider cannot get [25, 37]. The benefits in form of selective incentives are connected to the development process of the good and thus only accessible for the participants. Therefore, OSS contributions cannot be seen as pure public goods as these have significant private elements that evolved out of the ideology, which support the community [11].

In sum, the private-collective model of innovation combines the advantages of both private investment and collective action model. Table 1 compares the most important aspects of the three innovation models from an economic perspective and in relation to OSS development.

Table 1. Comparison of different aspects for the private, collective and private-collective innovation model (Source: adapted from Demil and Lecocq [38] and Schaarschmidt et al. [39])

	Private	Collective	Private-Collective
License	Proprietary	Open	Open
Copyright ownership	Company	Collective	Collective
Number of participating companies	One	None	One or more
Revenue stream	Direct	None	Indirect
Control intensity	High	Low	Low
Knowledge sharing intensity	Low	High	High

3 Method

3.1 Research Context

To find a relevant OSS project for our research, we have taken different aspects into account (e.g., size of the project, activity and continuity, company involvement, availability of a large set of data). Finally, we chose the LK project as our research context, which has served as an example for OSS in many previous studies (e.g., [40]).

The LK project was initialized by Linus Torvalds in 1991 and has been one of the most active OSS projects since its beginning. There are software releases every three month on average, which are possible because of the fast-moving development process and the broad foundation of contributors, ranging from hobbyists to companies. Thus, it involves more people than any other OSS project. The kernel itself makes up the core component of any Linux system and is used in operating systems for mobile devices right up to operating systems for supercomputers. Typically, a new release of the kernel comprises over 10,000 patches contributed by over 1,100 developers representing over 225 companies and is published under the GNU General Public Licence v2 [8].

Besides the fact that the LK is one of the largest cooperative software projects ever started, it has also an economic relevance, as many companies have business models that rely on the LK or on software working on top of the LK, respectively. Many of these companies do actively participate in the improvement of the kernel and thereby take effect on the orientation of the development. Very active companies in the kernel development, among others, are RedHat, Intel, IBM, Samsung, Google and Oracle [8].

3.2 Data Collection and Coding of Contributor Categories

To obtain the data needed for our research, we crawled the LK mailing list web archive[2]. We use mailing list data as it is suitable to calculate network positions that

[2] http://marc.info/?l=linux-kernel.

represent developers' influence [15]. The "linux-kernel"[3] mailing list has the purpose of discussing LK development topics as well as of reporting bugs. The observation period of the LK community ranges from 1996 (beginning of the web archive) to 2014.

We identified actors that occur multiple times on the list, for example with different email addresses, but identical sender names. We have mapped these to one person object related to the email address s/he has used when sending a message to the list for the first time.

The identified people interacting in the LK mailing list act partly on behalf of companies. To get a deeper understanding if the actors in the mailing list are affiliated with a firm we used the domain name of the email addresses to assign people to a contributor category. Developers sending messages from a domain indicating that the person is employed by the corresponding company are classified as employed contributors, whereas people using email addresses from public email providers such as yahoo.com were classified as hobbyists. Likewise, we identified developers with email addresses indicating universities and research institutions. Assigning LK actors to a contributor category was done in a semi-manual and semi-automated process in order to obtain a high accuracy of the attributions. Detailed information about the different contributor categories is provided in Table 2.

Table 2. Contributor categories

Contributor category	Description	Examples
Companies	People with email addresses from companies	intel.com, redhat.com
Hobbyists	People with private email addresses	gmail.com, yahoo.com
Universities	People from universities	columbia.edu, duke.edu
Research institutions	People from research institutions and public authorities (e.g., IEEE, government, military)	ieee.org, nasa.gov, army.mil

The cleaned dataset comprises 1,941,119 communication replies for the total time period with overall 86,509 contributors involved. The overall distribution of the contributor groups is made of 37.96 % of company developers, hobbyists represent 51.22 %, universities account for 9.65 % and research institutions make up 1.17 %. Descriptive information about the dataset is given in Figs. 1 and 2. Figure 1 shows the quantity of identified contributors per contributor group and year from 1996 to 2014. Figure 2 states the amount of messages sent per contributor group and year for the investigated period.

3.3 Social Network Analysis

A social network represents persons connected by edges. Social networks can represent friendship relationships, communication, interaction contacts or other types of social

[3] http://vger.kernel.org/vger-lists.html#linux-kernel.

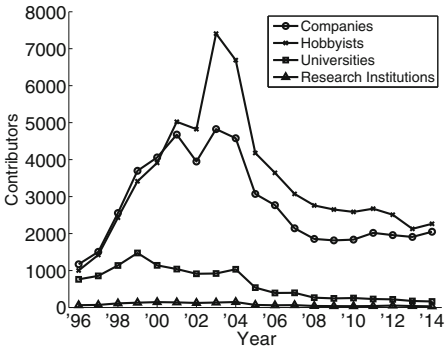


Fig. 1. Contributors per group and year

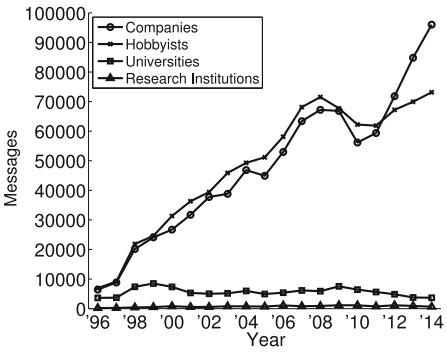


Fig. 2. Amount of messages sent per group and year

relationships. Social network datasets are widely used, not only in the area of social network analysis, but also in the areas of data mining, sociology, politics, economics and other fields [41].

In order to study the interactions of developers in the LK community, we perform an analysis of the LK mailing list's communication network. Communication within the Linux developer community can be modelled as a directed network, in which nodes are developers and directed edges (i.e., arcs) are a reply of one developer to another. In our dataset, we ignore all messages that are not replies to other developers. A relationship between the sender of a starter message and others does only emerge when one or more people reply to the starter message. We perform a structural analysis of this network to study the interplay of developers interacting.

The directed network of replies we consider is annotated with two additional metadata:

- For replies, the posting timestamp is known. This allows us to make a longitudinal analysis of the considered network statistics.
- For developers, we know their company, university or other affiliation, if any, allowing us to identify four categories of developers, as described in Sect. 3.2.

We perform social network analysis with Matlab and the KONECT Toolbox [42].

The contribution of one user in a directed social network can be used to measure both the activity and the importance of that user in the community. We achieve this by considering the following network-based measures, each of which is defined for individual nodes:

- The in-degree of a node equals the total number of replies received by a developer. The in-degree can thus be interpreted as a measure of importance of a developer.
- The out-degree of a node equals the total number of replies written by a developer, and can thus be interpreted as a measure of the activity of a developer.

- As a network-wide measure, we additionally define the Gini coefficient of the in-degree distribution [43], which denotes the inequality of the in-degrees. It is zero when all developers have equal in-degrees and one when a single developer received all replies. It can thus be interpreted as a measure of diversity of the community [44].

4 Results

4.1 Comparison of In-Degree and Out-Degree

In a first analysis, we compare the in-degree and the out-degree of all developers, i.e., the number of replies given vs. the number of replies received. Figure 3 shows the results of this analysis. We can observe that both measures are highly correlated – developers who receive many replies also write many replies. Thus, for the LK community the activity and the importance of developers correlate highly.

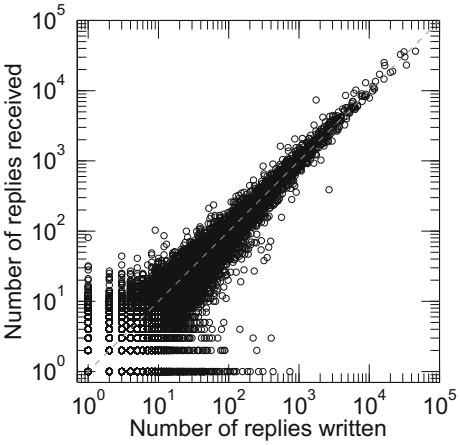


Fig. 3. Comparison of in-degree and out-degree

4.2 Comparison of Degree Per Group

In this analysis, we want to find out whether the developer-based measures of activity and importance vary from one group to another. We compute the distributions of out-degree and in-degree, for each group for the whole dataset aggregated over all years. The results are shown in Figs. 4 and 5.

The plots show that:

- The highest activity as measured by the out-degree is achieved by company developers, then hobbyists, and the lowest activity is given by developers from research institutions and universities.

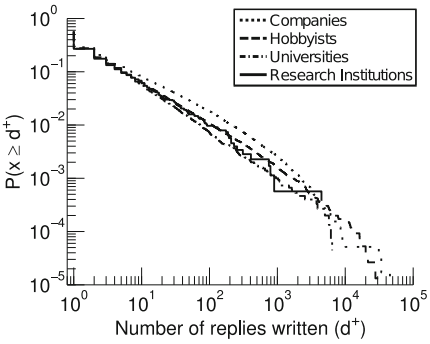


Fig. 4. Out-degree distribution

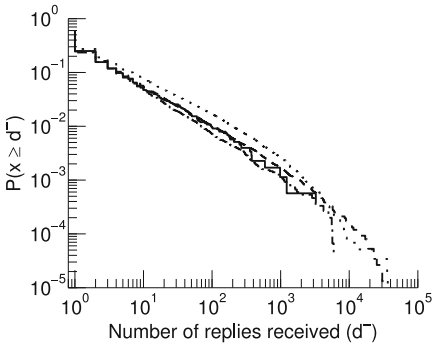


Fig. 5. In-degree distribution

- The measure of importance, the in-degree, correlates and shows the same pattern as for the activity: company developers have the most importance, then hobbyists, and finally developers from research institutions and universities.

These results are consistent with the observation that the measures of activity and importance correlate.

To verify the statistical significance of our results, we perform pairwise Mann–Whitney U tests, testing whether values of each statistic for one type of developer are statistically different from the values for another group. The group differences are statistically significant ($p < 0.05$; company developers vs. hobbyists: $p < 0.10$ for the out-degree), except for developers from companies vs. developers from research institutions and hobbyists vs. developers from research institutions for the in-degree and out-degree; developers from universities vs. developers from research institutions for the out-degree.

4.3 Longitudinal Analysis

In order to study the change of the community over time, we compute three group-wide measures of activity and importance for each individual year in the range 1996 to 2014.

- The average value of the out-degree and the in-degree of all developers in each group, restricted to all replies given and received, respectively during a given year.
- The Gini coefficient of the in-degree distribution of all developers of a given group, restricted to all replies received during a given year.

The results of the analysis are shown in Figs. 6, 7 and 8.

The average out-degree and in-degree (Figs. 6 and 7) show a consistent result with the degree distribution shown in Figs. 4 and 5. The average out-degree standing for activity of the developers of the different groups increases for the developers from companies as well as hobbyists over time and does not chance significantly for the developers of the other contributor categories. The measure of importance, the in-degree, shows a similar behavior. The values for the developers from companies and hobbyists increase and do not change significantly for the other types of developers.

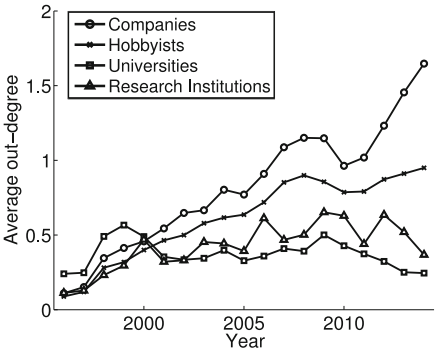


Fig. 6. Average out-degree

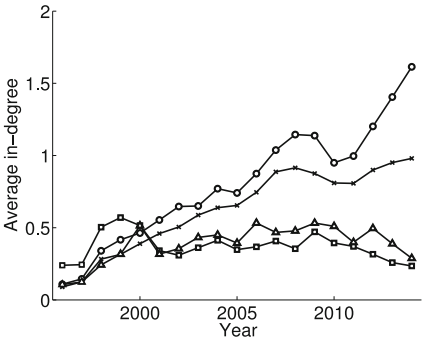


Fig. 7. Average in-degree

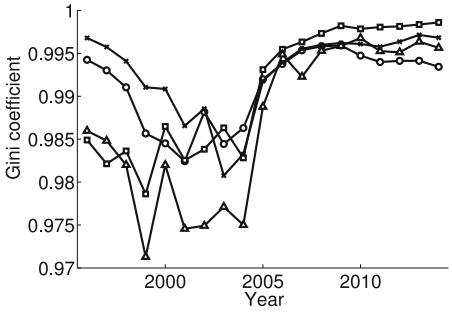


Fig. 8. Gini coefficient

The network wide measure Gini coefficient (Fig. 8) decreases slightly for developers from companies in the last years. The high fluctuation of the Gini coefficient for universities and research institutions is related to the small group sizes.

Across all developer groups and times, the Gini coefficient is very near to one (>98 %, up to fluctuations). This value is higher than in the large majority of social networks [43], indicating that the importance in the community is concentrated in a very small number of actors when compared to other typical social networks.

5 Discussion and Conclusion

5.1 Discussion

This study provides results of an activity analysis of different contributor groups in the LK development from 1996 to 2014. The aim of this study was to investigate how different contributor groups associated with public and increasing private interests interact in an OSS development project. To achieve this goal, we analyze developers active in the LK development community from a social network perspective, as the interaction between members of a software development community reflects the structure of their collaboration.

The first result of our analysis shows that the out-degree, as a measure for activity, and the in-degree, as a measure of importance, correlate highly both in general and for each contributor group individually. Thus, developers who write many replies to the mailing list also receive many replies. This phenomenon in the LK mailing list differs from forum communication, where a variety of user roles with different communication patterns can be identified [45]. Both the highest activity and importance, is achieved by company developers followed by hobbyists. The lowest activity and importance can be attributed to developers from research institutions and universities. Connecting these results to the amount of contributors per contributor category (Fig. 1) it can be seen that although the amount of developers from companies is less than the amount of hobbyists the impact made by employed developers on the LK community is larger. The mentioned impact is expressed by the amount of messages sent per contributor group and year (Fig. 2) as well as stated by the measures of activity (Fig. 6) and importance (Fig. 7), clearly seen especially from 2010 on.

The second result of our analysis shows that the Gini coefficient, as a network wide measure, decreases in recent years for developers from companies and remains constant for the other groups. Although the decrease is small, this can be seen as a tendency that the importance in the community for the group of developers from companies is distributed to more actors. When considering the early years of the LK development (from 1996 until 2000) it can be seen that university members were the most active and important force in the project (Figs. 6 and 7). From 2001 the activity and importance of hobbyists and companies increases alike. Although the amount of company contributors remains relatively stable for the period starting from 2010 until 2014 (Fig. 1) the activity (Figs. 2 and 6) and importance (Fig. 7) in that time increases sharply.

5.2 Conclusion

The aforementioned observations help to answer the research question of how different contributor groups associated with public and increasing private interests interact in an OSS development project, here the LK development. In the beginning, the LK project was driven by intrinsic motivated enthusiasts who are hobbyists and university members (e.g., students) and are associated with pure collective interests. As software and services that were built on top of the LK got more and more influence, the participation of firms, for example, motivated by offering complementary services (see Sect. 2.1), increased. As a consequence the private interests in this project increased, too, especially from 2010 on with the diffusion of mobile devices powered by Android [46], which uses the LK as foundation of the operating system.

Summarizing the results of this study, it can be concluded that the balance between private and collective contributors in the LK development seems to be changing to an open source project that is mostly developed jointly by private companies. These firms need the kernel for their products and services. Thus, the LK project is no longer just an open source alternative for hobbyist to develop open source software for the reason not be locked in by proprietary software. The advantages for the participating companies outweigh the drawbacks in terms of collective copyright ownership and less control in the project (see also Sect. 2.2, Table 1) as the LK community can be utilized to complement their own resource base for innovations.

5.3 Implications for Research

Our findings can be classified into the context of private-collective innovation. The LK development project is an outstanding OSS project. The engagement of companies has increased in the last years, as more and more firms have business models that rely on the kernel. Although companies cannot dictate what the community should do, they can in a way influence the trajectory of the project by assigning employed developers to the project [10, 47]. As the results of our study for the LK project show, employed developers can take key positions in a community due to the intensity of the commitment, expressed by activity and importance, for the community. With this in mind, future research should more thoroughly discuss the nature and structure of firm presence in OSS development. The majority of early research on OSS somehow neglected firm presence and centered on developer motivation while later research discussed emerging OSS business models (e.g., [1]). However, our study calls for more longitudinal studies on firm presence in OSS as (1) firm engagement varies over time and (2) former hobbyists might transform into employed developers (the latter was no focus of this study).

5.4 Limitations and Suggestions for Future Research

Our research has some limitations that have to be considered when utilizing our study's outcomes. The categorization of LK mailing list actors by the hostname of their email addresses into four contributor groups is an approximate but sufficient classification. It is known that there are developers that do not use their company email address while contributing and actors may do personal work out of the office [8]. We mapped developers that occur more than once on the LK mailing list to one person object related to the email address s/he has used when sending a message to the list for the first time. Multiple occurrences happen if the message sender uses different email addresses over time, but the same sender name (e.g., because of company changes). We have not considered these dynamics in our analysis. Furthermore, it has to be considered that the LK project is a unique OSS project, so that our conclusions cannot directly be transferred to other OSS projects where firm-sponsored developers are involved. Further research can consider to investigate the different types and content of the interaction as well as the aforementioned dynamics of developers or compare the multi-vendor project LK to single-vendor OSS projects.

References

1. Bonaccorsi, A., Giannangeli, S., Rossi, C.: Entry strategies under competing standards: hybrid business models in the open source software industry. Manage. Sci. **52**, 1085–1098 (2006)
2. Capra, E., Francalanci, C., Merlo, F., Rossi-Lamastra, C.: Firms' involvement in open source projects: a trade-off between software structural quality and popularity. J. Syst. Softw. **84**, 144–161 (2011)

3. Riehle, D.: The single-vendor commercial open source business model. Inf. Syst. e-Business Manage. **10**, 5–17 (2012)
4. Guignard, C.: Ten facts and figures about open source software. http://blog.nxcgroup.com/2013/ten-facts-and-figures-about-open-source-software/
5. IDC: Worldwide Open Source Software 2009–2013 Forecast (2009)
6. Grand, S., Von Krogh, G., Leonard, D., Swap, W.: Resource allocation beyond firm boundaries: a multi-level model for open source innovation. Long Range Plan. **37**, 591–610 (2004)
7. Schaarschmidt, M., Von Kortzfleisch, H.: Divide et Impera! The role of firms in large open source software consortia. In: Proceedings of the 15th Americas Conference on Information Systems (AMCIS), San Francisco, USA (2009)
8. Corbet, J., Kroah-Hartman, G., McPherson, A.: Linux Kernel Development. How Fast It is Going, Who is Doing It, What They are Doing, and Who is Sponsoring It. Linux Foundation, San Francisco (2013)
9. Teigland, R., Di Gangi, P.M., Flåten, B.-T., Giovacchini, E., Pastorino, N.: Balancing on a tightrope: managing the boundaries of a firm-sponsored OSS community and its impact on innovation and absorptive capacity. Inf. Organ. **24**, 25–47 (2014)
10. Schaarschmidt, M., Von Kortzfleisch, H.: Firms' resource deployment and project leadership in open source software development. Int. J. Innov. Technol. Manage. **12**, 1–19 (2015)
11. Von Hippel, E., Von Krogh, G.: Open source software and the "private-collective" innovation model: issues for organization science. Organ. Sci. **14**, 209–223 (2003)
12. Stürmer, M., Späth, S., Von Krogh, G.: Extending private-collective innovation: a case study. R&D Manage. **39**, 170–191 (2009)
13. Alexy, O., Reitzig, M.: Private-collective innovation, competition, and firms' counterintuitive appropriation strategies. Res. Policy **42**, 895–913 (2013)
14. Riehle, D., Riemer, P., Kolassa, C., Schmidt, M.: Paid vs. volunteer work in open source. In: Proceedings of the 47th Hawaii International Conference on System Science (HICSS), Hilton Waikoloa, Big Island (2014)
15. Dahlander, L., Wallin, M.W.: A man on the inside: unlocking communities as complementary assets. Res. Policy **35**, 1243–1259 (2006)
16. Open Source Initiative: The open source definition. http://opensource.org/osd
17. Preece, J.: Online Communities. Designing Usability and Supporting Sociability. Wiley, New York (2000)
18. Hars, A., Ou, S.: Working for free? Motivations for participating in open-source projects. Int. J. Electron. Comm. **6**, 25–39 (2002)
19. Lakhani, K.R., Wolf, R.: Why hackers do what they do: understanding motivation and effort in free/open source software projects. In: Feller, J., Fitzgerald, B., Hissam, S.A., Lakhani, K. R. (eds.) Perspectives on Free and Open Source Software, pp. 3–22. MIT Press, Cambridge (2005)
20. Roberts, J.A., Hann, I.-H., Slaughter, S.A.: Understanding the motivations, participation, and performance of open source software developers: a longitudinal study of the apache projects. Manage. Sci. **52**, 984–999 (2006)
21. Wu, C.-G., Gerlach, J.H., Young, C.E.: An empirical analysis of open source software developers? Motivations and continuance intentions. Inf. Manage. **44**, 253–262 (2007)
22. Ryan, R.M., Deci, E.L.: Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp. Educ. Psychol. **25**, 54–67 (2000)
23. Baytiyeh, H., Pfaffman, J.: Open source software: a community of altruists. Comput. Hum. Behav. **26**, 1345–1354 (2010)
24. Stewart, K., Gosain, S.: The impact of ideology on effectiveness in open source software development teams. MIS Q. **30**, 291–314 (2006)

25. Lerner, J., Tirole, J.: Some simple economics of open source. J. Ind. Econ. **50**, 197–234 (2002)
26. Dahlander, L., Magnusson, M.: How do firms make use of open source communities? Long Range Plan. **41**, 629–649 (2008)
27. Andersen-Gott, M., Ghinea, G., Bygstad, B.: Why do commercial companies contribute to open source software? Int. J. Inf. Manage. **32**, 106–117 (2012)
28. Fitzgerald, B.: The transformation of open source software. MIS Q. **30**, 587–598 (2006)
29. Dahlander, L.: Appropriation and appropriability in open source software. Int. J. Innov. Manage. **9**, 259–285 (2005)
30. Riehle, D.: The economic motivation of open source software: stakeholder perspectives. Computer **40**, 25–32 (2007)
31. Henkel, J.: Open source software form commercial firms: tools, complements, and collective invention. Zeitschrift für Betriebswirtschaft **4**, 1–23 (2004)
32. Hawkins, R.E.: The economics of open source software for a competitive firm. NETNOMICS Econ. Res. Electron. Netw. **6**, 103–117 (2004)
33. Ågerfalk, P.J., Fitzgerald, B.: Outsourcing to an unknown workforce: exploring opensourcing as a global sourcing strategy. MIS Q. **32**, 385–409 (2008)
34. Liebeskind, J.P.: Knowledge, strategy, and the theory of the firm. Strat. Manage. J. **17**, 93–107 (1996)
35. Olsen, M.: The Logic of Collective Action. Harvard University Press, Cambridge (1967)
36. Von Krogh, G.: The communal resource and information systems. J. Strat. Inf. Syst. **11**, 85–107 (2002)
37. Raymond, E.S.: The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media Inc., Sebastopol (2001)
38. Demil, B., Lecocq, X.: Neither market nor hierarchy nor network: the emergence of bazaar governance. Organ. Stud. **27**, 1447–1466 (2006)
39. Schaarschmidt, M., Walsh, G., MacCormack, A., Von Kortzfleisch, H.: A problem-solving perspective on governance and product design in open source software projects: conceptual issues and exploratory evidence. In: Proceedings of the International Conference on Information Systems (ICIS), Research-in-Progress, Milan, Italy (2013)
40. Lee, G.K., Cole, R.E.: From a firm-based to a community-based model of knowledge creation: the case of the Linux kernel development. Organ. Sci. **14**, 633–649 (2003)
41. Wasserman, S., Faust, K.: Social Network Analysis. Methods and Applications. Cambridge University Press, Cambridge (1994)
42. Kunegis, J.: KONECT - the Koblenz network collection. In: Proceedings of the 22nd International Conference on World Wide Web Companion, Rio de Janeiro, Brazil (2013)
43. Kunegis, J., Preusse, J.: Fairness on the web: alternatives to the power law. In: Proceedings of the Web Science Conference 2012, Evanston, IL, USA (2012)
44. Kunegis, J., Sizov, S., Schwagereit, F., Fay, D.: Diversity dynamics in online networks. In: Proceedings of the 23rd Conference on Hypertext and Social Media, Milwaukee, WI, USA (2012)
45. Chan, J., Hayes, C., Daly, E.: Decomposing discussion forums using common user roles. In: Proceedings of WebSci 2010: Extending the Frontiers of Society On-Line, Raleigh, North Carolina (2010)
46. Google: Android @ i/o: the playground is open. Official Google blog. http://googleblog.blogspot.de/2012/06/android-io-playground-is-open.html
47. Schaarschmidt, M., Walsh, G., Von Kortzfleisch, H.: How do firms influence open source software communities? A framework and empirical analysis of different governance modes. Inf. Organ. **25**, 99–114 (2015)

Towards a Set of Capabilities for Orchestrating IT-Outsourcing in the Retained Organizations

Bas Kleinveld[1,2(✉)] and Marijn Janssen[2]

[1] The Orchestrate Group B.V., Lichttoren 32, 5611 BJ Eindhoven
The Netherlands
bas.kleinveld@orchestrate-group.com
[2] Faculty of Technology, Policy and Management,
Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands
M.F.W.H.A.Janssen@tudelft.nl

Abstract. Managing outsourcing processes requires a retained organization to orchestrate an organization's IT functions into a concerted whole. Nowadays multiple outsourcing vendors need to be managed. The purpose of this research is to identify capabilities affecting orchestration of outsourced IT functions. An existing framework for outsourcing capabilities is used as a starting point. Due to the scarcity of research a qualitative research approach based on two case studies is taken. The findings shows that the core IS capabilities as found in the literature are found to be too abstract, ambiguous and that several essential capabilities are missed. A framework containing a refined and extended set of capabilities is derived. Four new capabilities were found including: demand, financial, (service) delivery and service portfolio management.

Keywords: Capabilities · IT outsourcing · Multi-vendor outsourcing · Capability · Orchestration · Retained organization

1 Introduction

Many organizations have outsourced (part of) their IT functions to one or more outsourcing vendors. The IT outsourcing and services offering have changed over time. Due to technology developments such as cloud and market developments companies do not rely on a single vendor any more. Many organizations struggle with their sourcing strategies; retained capabilities and the way function to orchestrate the sourcing providers should be embedded in the organization. In practice, existing capabilities fall short and retained organizations should consider improving their capability in orchestrating sourced IT functions to achieve their objectives. This paper investigates capabilities needed by retained IT organizations for orchestrating sourced IT functions.

When outsourcing IT to other parties the IT still need to be managed by the outsourcing party. For this purpose the client needs a *retained organization*, which orchestrates the various outsourcing vendors. Troost (2009) defined the retained organization as the mediator between the client organization and the internal and

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 314–325, 2015.
DOI: 10.1007/978-3-319-25013-7_25

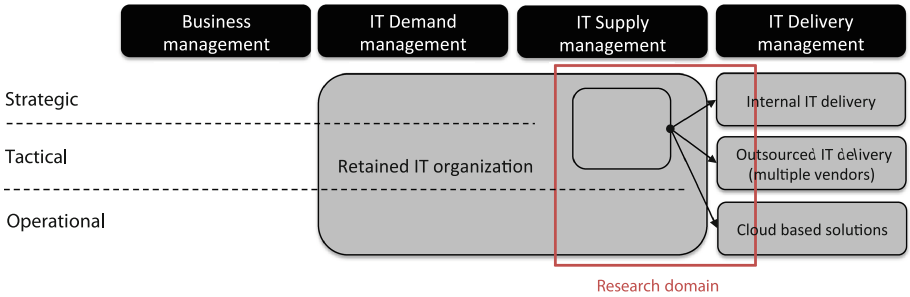


Fig. 1. Research domain

external service suppliers when outsourcing activities. The Fig. 1 shows the retained organization, which is matching the demands from the business (IT demand management) and with the delivery of IT services by either internal or external delivery partners (IT supply management).

The retained organization needs capabilities for managing the interdependencies with the multiple outsourcing vendors. In this paper we focus on strategic and tactical capabilities. The main reason for focusing on strategic and tactical capabilities is that operational capabilities are associated with delivery functions performed mostly by suppliers or operational delivery teams, whereas the strategic and tactical aspects of the capabilities are relatively unknown. Lots of non-academic literature on management and governance frameworks is available, but limited scientific literature was identified investigating capabilities needed for orchestrating IT functions. The *goal* of this research is to identify capabilities affecting the orchestration of sourced IT functions.

This paper is structured in sections as follows. First we present the reviewed literature on the capabilities of retained organizations and orchestrating the sourced IT-function. Next the research approach is presented followed by the description of case studies. The case studies include the analysis of the orchestration function, as well as the identified capabilities. These capabilities are elaborated on and further discussed. Finally, conclusions are drawn.

2 Literature Background

A study performed by Gartner (2003), shows that enterprise organizations have some key challenges across IT within their organizations. In planning IT outsourcing, organizations seem to have difficulties in deciding which specific roles and capabilities have to be retained in the organization in order to mitigate risk, manage the IT supplier and manage the business demand appropriately Joha (2003). Organizations have a set of capabilities to remain competitive and conduct their activities. Willcocks et al. (2006) defined a capability as a distinctive set of human resource-based skills, attitudes, motivations and behaviors that contribute to achieving specific activities and influencing business performance. In 1998, Feeny and Willcocks developed an IT governance and management framework out of two streams of research. The first

stream concerned characteristics of high performers in the Information Technology (IT) function; the second examined the retained capabilities needed to run effective IT outsourcing deals. Feeny and Willcocks (1998) synthesized the findings into an IT governance and management framework suggesting four tasks of the future IT function and the nine core capabilities organizations need to retain in-house.

In detailed case research into major IT outsourcing arrangements Lacity and Willcocks (2000) found that the relationship dimension between client and suppliers a critical, but also complex to manage. According to Willcocks et al. (2006) governance has a central task to devise organizational arrangements (structures, processes and staffing) to successfully manage the interdependencies, and ensure that the IT function delivers value for money. The model from Willcocks et al. (2006) presents a number of serious human resource challenges. It requires high performers in each role. Furthermore, in contrast to the more traditional skills found in IT functions there needs to be a much greater emphasis on business skills and orientation in all but the two very technical capabilities, i.e. architecture planning and making technology work. There is a significantly increased requirement for 'soft' interpersonal skills across all roles, all roles demand high performers, and each role requires a specific set of people behaviors, characteristics and skills (Willcocks et al. 2006). Managing the outsourcing processes required a retained organization to orchestrate the IT functions in a concerted whole. Empirical research remains scarce and there is little known about the underlying theories and management principles. One exception is Plugge and Janssen (2014) who investigated multi-vendor governance by focusing on how resources are coordinated and which resources the organizations are dependent on.

When orchestrating a portfolio of capabilities some capabilities will be executed adequately, others poorly; but a few must be superior to the competition if the business is to sustain a market position that is valuable and difficult to match (Day 1994). Managers must become thought leaders and strategic architects in determining the company's core functions, building on an in-depth understanding of the potential competitiveness and value contributions of external providers (Spiller et al. 2014).

In multivendor outsourcing, vendors experience a strong degree of distrust between each other, due to lack of unclear strategic roles and responsibilities and lack of structure (e.g. meetings, forecasts). In particular, the vendors' inability or unwillingness to cooperate in delivering IT services to the client affected their relationship negatively. The various roles, activities and responsibilities of each party are often not clear (Plugge and Bouwman 2013). Given the strategic nature of their IT-outsourcing goals, it is important for outsourcing-driven and outsourcing-centric customers to make relational investments in their partnerships with vendors (Mehta and Mehta 2010). The coordination of IT activities that reflects the relationship between the client and the vendors, however, reveals that no particular coordination mechanism is used. Based on analyses by Plugge and Janssen (2014) research demonstrates that clients and vendors need to develop and implement clear governance structures and mechanisms to coordinate the delivery of IT services successfully. The executive management of both client and vendors has to implement these governance structures to be able to proactively monitor changes in a multivendor landscape.

Governing multivendor outsourcing arrangement is a continuous process that requires regular management attention (Plugge and Janssen 2014). Managing complex

vendor relationships and contracts requires levels of project management skills typically expected of category managers. This creates significant opportunities to lead increasingly complex sourcing relationships and manage these external relations to maximize value. According to Coltman and Devinney (2013) the need to match supply with demand has gained a prominent position in the service operations and the wider operations management literature. However, the role that managers play in the alignment process has been rarely investigated. Several companies have recognized the need to improve the end-to-end orchestration of their value chains (Spiller et al. 2014).

The findings of a study by Plugge et al. (2013) demonstrates that outsourcing capabilities and organizational dimensions are perceived to be critical factors in achieving quality performance, and that a fit between them is paramount. The outcome of this study demonstrates that the outsourcing experts perceived that the client's need was less important in influencing organizational structure (Plugge et al. 2013). The results suggest that monitoring and assessing changing client circumstances regularly is a prerequisite for providers to be an agile organization. Consequently, they must be willing and able to adapt their outsourcing capabilities and organizational structure to achieve high quality performance and thus to remain competitive.

3 Research Approach

The goal of this research is to identify capabilities affecting the orchestration of sourced IT functions. This research focuses on situations in which services are continuously delivered. The main reason for this focus is that orchestrating continuous service delivery is complex in nature, since it is related to multiple internal and external suppliers working together. Case study research is employed to gain insight into the capabilities desired and the actual capabilities an organization has to orchestrate. The case studies were investigated by reading reports and conducting interviews. In total two case studies were investigated meeting the following requirements. The companies should be an international company with more than 20,000 employees. The organization should have a mix of internal delivery teams and external multiple vendors and the IT budget should exceed 100 million euro's.

Semi-structured interviews have been conducted on the required core capabilities in orchestrating sourced IT functions will be held. The interviews were conducted between January and May 2015, with 6 senior managers and took 60 to 90 min and are important since interviews enable the collection of in-depth information from senior managers. The interview scheme is included as appendix to this paper. Interviews were conducted with the service delivery manager, IT director Applications, EMEA IT manager, program director, director of sourcing, and IT director Platforms.

4 Case Studies

The case studies help us to gain results gain in-depth insight in the nature of orchestration function in IT organizations and the capabilities needed and actual capabilities for orchestration.

4.1 Case 1

The company is one of Europe's largest electricity companies with approximately 31,000 employees. Operations are conducted in different European countries with revenues exceeding 15 billion euro in 2013. The business activities are divided into several Business Units.

The IT organization is divided in a demand and supply management structure. How the demand and supply interact with one another is a central factor for the success of the IT strategy and orchestration function. The IT organization is based on a federated organizational model, where all corporate resources are allocated to IT. The company has over 2,500 applications, of which around 1,200 applications are supported (and financed) by IT. IT supply and delivery management is divided in 3 main IT delivery pillars: IT infrastructure, and 2 application solutions groups.

4.2 Case 2

This company is a leading global manufacturing company based in The Netherlands. The company supplies industries and consumers worldwide with innovative products. Its portfolio includes well-known international brands. The company is consistently ranked as one of the leaders in the area of sustainability. The company has global operations with more than 50,000 employees. The revenue in 2013 exceeded 15 billion euro. The company is organized in three main business units.

The IT organization is divided in a demand and supply management structure. Demand and Supply Management manages the interface between business and IT, maintaining the balance of demand and supply of IT services. IT supply is divided in 3 main delivery pillars: Infrastructure services, and two application services groups.

4.3 Capabilities in the Orchestration Function

When analyzing the orchestration function of both cases, the companies struggle with their sourcing strategies, related capabilities and the way orchestration function should be embedded in the organization. In interviews and documentation, core capabilities to perform the orchestration function were identified. In both cases the core capabilities are primarily performed by external resources. The main reason is the lack of required skills within the organization, and lack of available resources to perform the activities. As a consequence orchestration function in organizations did not have a high maturity level. Further research in this domain is recommended (Table 1).

Table 1. Overview of capabilities

Type of capabilities	Case: Company X	Case: Company Y
IT-Architecture	• IT architecture consists of enterprise architecture, solution architecture and domain knowledge/ architecture • Guarding existing templates/systems while keeping-up the pace with new technologies	• IT architecture is considered enterprise architecture, solution architecture, and to a limited extend domain and technical architecture • Although architecture is considered a core capability, it can be sourced from externals in certain situations
Contract management and sourcing	• Managing contracts between key suppliers and the business • Not implemented at Company X, but is considered key	• Managing (outsourced) contracts and vendors • Contract management is performed by Sourcing (purchasing), not by IT
Supplier and vendor management	• Managing supplier performance • Managing the vendors and the related IT services • Supplier management is not implemented during times of this research	• Managing suppliers on the operational performance, KPI management and operational budget Contract management is performed by IT
Team and relationship management	• Managing relationships, stakeholder and people is a capability often underestimated, it remains a people business • Right team balance between performance, capabilities, internal and external resources	• Close relationships with the business on their needs and requirements will improve the orchestration function related to sourced IT functions
Content knowledge related to business domains and (ERP) systems	• Content knowledge is understanding of the business processes and map to functionality and IT solutions • Performed by many externals	• Content knowledge is understanding of the business processes and map to functionality of the applications in use (strategic applications) • Performed by many externals, although there is a clear strategy to replace externals on this capability
Demand management	• Stakeholder management • Understanding of business needs • Managing the business demand for IT	• Managing the demand of IT needs from the business • Mapping the needs to right IT groups

(*Continued*)

Table 1. (*Continued*)

Type of capabilities	Case: Company X	Case: Company Y
Service delivery management	• Managing the IT solutions operational delivery (on both internal and external side) on a day-to-day basis	• Managing the operational delivery on a day-to-day basis including quality assurance
Financial management	• Mapping the costs to activities, matching this to the actual and forecasted costs made by suppliers	• Charge out model for projects and support services • Manage supplier statement of works and service level agreements to invoicing and financial bookings • Although this is considered a core capability, this activity is not optimal performed
Service portfolio management	• Not indicated as capability by interviewees	• Defining the services to be delivered to the business • IT departments will act more as service providers to their internal customers • Ability to understand what to deliver to internal customers

4.4 The Orchestration Function Analyzed

During interviews, the interviewees were asked which capabilities influenced their orchestration function in IT. These capabilities were collected and analyzed using a data matrix. Table 2 visualizes the capabilities provided by the interviewees from the case studies, including additional capabilities. Some capabilities in the case studies have been confirmed during the interviews, although there is still a gap between case study capabilities and the capabilities identified in literature.

In our research we found 4 capabilities in addition to the capabilities suggested by Feeny and Willcocks (1998) and Willcocks et al. (2006). Demand management, financial management, (service) delivery management and service portfolio management were found as capabilities, which were mentioned by the interviewees as shown at the bottom of the table. Several capabilities have been integrated to a single capability based on the interview outcomes. Making technology work and architecture planning are IT-architecture, while informed buying, contract facilitation, contract monitoring and vendor development are considered supplier management, contract management and sourcing. Relationship management has been extended to team and relationship management. The capabilities from the case studies and indicated by interviewees are described hereafter.

4.5 IT-Architecture

IT-Architecture was mentioned by almost all interviewees consists of enterprise architecture, solution architecture, and to a limited extend domain and technical architecture. Interviewees indicate that although architecture is considered a core capability, it can be sourced from externals in certain situations. An interviewee quotes the importance of architecture as: "There are always excuses for not using the defined architectural standards. Finding reasons for exceptions creates complexity and issues in organizing IT services. The challenge is that IT delivery functions need to keep up the pace with new technologies while keeping alive the legacy technologies and systems".

4.6 Supplier Management, Contract Management and Sourcing

Supplier management, contract management and sourcing are recognized as a core capability by all interviewees to perform the orchestration function of IT. Although this is recognized, several interviewees see developments in this domain. It is recognized as core capability while interviewees indicate that supplier and contract management is a domain for improvement.

4.7 Team and Relationship Management

Interviewees recognize the need for relationship management and good team management. Managing relationships, stakeholder and people is a capability often underestimated, as all performed activities is still about people. Close relationships with the business on their needs and requirements will improve the orchestration function related to sourced IT functions (interviewee). "At the moment we have a balance and right mix of capabilities. Externals contribute to the right mix of this balance in team management".

4.8 Content/Technical Knowledge Related to Business Domains/Functions

Many interviewees indicate that content and technical knowledge related to business domains and functions is a key capability in IT departments, and a key capability in orchestrating sourced IT functions. Content knowledge relates to understanding of the customer's business and translate this to IT solutions, but also related to the understanding of processes and company dynamics. In addition resources with technical knowledge are needed to be able to judge supplier and internal delivery performance, related to their objectives. Content and technical knowledge related to business functions and domains can relate back to architecture. Some capabilities that have been indicated by interviewees do not match the capabilities from literature and the case studies. Service integration management was a capability from the case studies that was not mentioned by interviewees.

Table 2. Identified capabilities

Identified Capabilities		TOTAL	case 1			case 2		
			Customer organization P1	Customer organization P2	Customer organization P3	Customer organization P4	Customer organization P5	Customer organization P6
Capability from literature (Feeny & Willcocks, 1998)								
1. Architecture planning	IT-Architecture (4.5)	**5**	x		x	x	x	x
2. Making technology work	IT-Architecture (4.5)	**5**	x		x	x	x	x
3. Informed buying	Contract management and sourcing (4.6)	**6**	x	x	x	x	x	x
4. Contract facilitation	Contract management and sourcing (4.6)	**6**	x	x	x	x	x	x
5. Contract monitoring	Contract management and sourcing (4.6)	**6**	x	x	x	x	x	x
6. Vendor development	Supplier & vendor management (4.6)	**4**	x	x		x	x	
7. Relationship building	Team & relationship management (4.7)	**4**	x	x	x	x		
8. Business system thinking	Content & technical knowledge related to business functions (4.8)	**5**		x	x	x	x	x
Capabilities not found in literature								
	Demand management (4.9)	**3**				x	x	x
	Financial management (4.10)	**2**	x		x			
	(Service) delivery management (4.11)	**3**	x	x	x			
	Service portfolio management (4.12)	**1**	x					

4.9 Demand Management

Demand management is not a capability from literature, but indicated to be important from the cases. Managing the business demand has been indicated as a capability that influences the orchestration function by half of the interviewees. Customer intimacy and business information management are given as areas that are important for demand management.

4.10 Financial Management

Almost all interviewees have indicated that cost optimization if an important driver in their objectives and goals of IT departments. In the findings that influence capabilities, interviewees mentioned cost optimization as important issue. Although this is an important issue, financial management of their IT functions is only recognized by a number of interviewees, Customer organizations should consider financial management as capability in orchestrating sourced IT functions related to their financial objectives.

4.11 (Service) delivery management

The delivery of IT services to the business is considered a capability to organizations, as interviewees indicate that it will become even more important. "When outsourcing more services, the capability of delivery management changes. We've experiences this and now adapt the philosophy: eyes on, hands off", according to an interviewee. Mechanisms that can improve delivery functions mentioned across several interviewees is to introduce shared KPI's across multiple IT towers in delivery and more output measurement rather then activity management. Focus on end result rather then activity.

4.12 Service Portfolio Management

One firm believes that service portfolio management will become a more important capability now and in the future, as IT departments will act more as service providers to their internal customers. Interviewee quotes: "This is a core capability organizations should have to be able to understand what they deliver to their internal customers". Cloud solutions drives the business to buy services directly from the market, while the value for IT departments will decrease if they don't deliver services to their internal customers, rather then resources.

5 Discussion and Conclusions

In the two case studies the organizational members are aware of the importance of orchestrating sourced IT functions and the need to build the right capabilities. Although they are aware of which capabilities are needed, the existing capabilities fall short. In particular in the domains as contract, vendor and supplier management. Retained

organizations should consider financial management as capability in orchestrating sourced IT functions related to achieving their financial objectives. Many interviewees indicated that content and technical knowledge related to business functions is a key capability for their IT department, and a key capability in orchestrating sourced IT functions. Two interviewees indicated that although architecture is considered a core capability, it can be sourced from externals in certain situations.

Due to lack of required skills, resources and capabilities, the orchestration func-tion in organizations does not have the desired maturity level. When analyzing the orchestration function of the cases, the companies struggle with developing their orchestration capabilities and the way orchestration function should be organized in their organizations. In the interviews, core capabilities to perform the orchestration function are clearly identified, but these capabilities are often performed by external people or by persons not equipped to perform the role. The main reason for this is the lack of people having the required skills within the organization, and lack of people to perform the activities. The orchestration function is identified as an abstract capability consisting of a subset of capabilities that IT organizations should retain after outsourcing services to external providers.

When comparing to the findings of the case studies with the literature a clear gap can be seen. Feeny and Willcocks synthesized the findings into an IT governance and management framework suggesting four tasks of the future IT function and the nine core capabilities organizations need to retain in-house. The model from Feeny and Willcocks (1998) presents a number of serious human resource challenges. However literature limited attention is paid to invest in the relationship, while the cases show that relationship and team management are an essential capability for orchestration. Furthermore, in contrast to the more traditional skills found in IT functions there needs to be a much greater emphasis on business skills and orientation in all but the two very technical roles. Although the four tasks and nine capabilities (Feeny and Willcocks, 1998) show the need of these, the cases show that the capabilities identi-fied need to be practically applicable, and some of the capabilities are formulated at a too high level of abstraction for practical applicability in organizations. In our re-search we found 4 capabilities to the capabilities of Feeny and Willcocks (1998), which are demand management, financial management, (service) delivery manage-ment and service portfolio management. In our research we extended the capability relationship management with team and relationship management.

References

Coltman, T., Devinney, T.M.: Modeling the operational capabilities for customized and commoditized services. J. Oper. Manag. **31**, 555–566 (2013)

Day, G.S.: The capabilities of market-driven organizations. J. Mark. **58**(10), 49–63 (1994)

Feeny, D.F., Willcocks, L.P.: Core IS capabilities for exploiting information technology. Sloan Manag. Rev. **39**(3), 9–21 (1998)

Gartner Executive Report Series, IT Spending: How Do You Stack Up (2003). Retrieved from the Gartner Inc. http://www.gartner.com

Gartner predicts limited IT outsourcing growth and increased volatility (2013). http://www.cio.com/article/737472/Gartner_Predicts_Limited_IT_Outsourcing_Growth_and_Increased_Volatility

Joha, A.: The retained organization after IT outsourcing, the design of its organizational structure. Master Thesis, Delft University of Technology, Delft, The Netherlands (2003)

Lacity, M., Willcocks, L.: IT outsourcing relationships: a stakeholder perspective. In: Zmud, R. (ed.) Framing the Domains of IT Management Research. Glimpsing the Future Through the Past. Pinnaflex Educational Resources, Cincinnati (2000)

Mehta, N., Mehta, A.: It takes two to tango: how relational investments improve IT outsourcing partnerships. Commun. ACM **53–2**, 160–164 (2010)

Plugge, A.G., Bouwman, H.: Fit between sourcing capabilities and organizational structure on IT outsourcing performance. Prod. Plan. Control **24–4**(5), 375–387 (2013)

Plugge, A.G., Janssen, M.: Governance of multivendor outsourcing arrangements: a coordination and resource dependency view. Doctoral dissertation, Delft University of Technology, Delft, The Netherlands (2014)

Plugge, A.G., Bouwman, W.A.G.A., Molina-Castillo, F.J.: Outsourcing capabilities organizational structure and performance quality monitoring: towards a fit model. Inf. Manag. **50**(6), 275–284 (2013)

Spiller, P., Reinecke, N., Ungerman, D., Teixeira, H.: New role for the CPO: orchestrating the end-to-end value chain. Supply Chain Manag. Rev. **18**, 27–33 (2014)

Strategic Road Map for Outsourcing Competencies. Gartner research G00250619 (2013). Retrieved from the Gartner Inc. www.gartner.com

The Reality of IS Lite. Gartner research G00117022 (2003). Retrieved from the Gartner Inc. http://www.gartner.com

Troost, M.: Sourcing the retained organization in IT outsourcing. Master Thesis, Delft University of Technology, Delft, the Netherlands (2009)

Willcocks, L., Fenny, D., Olsen, N.: IS Capabilities: Feeny-Willcocks IT governance and management framework revisited. Eur. Manag. J. **24**(1), 28–37 (2006)

Why Do Small and Medium-Size Freemium Game Developers Use Game Analytics?

Antti Koskenvoima and Matti Mäntymäki(✉)

Turku School of Economics, Rehtorinpellonkatu 3, 20500 Turku, Finland
{antti.koskenvoima,matti.mantymaki}@utu.fi

Abstract. The increased use of the freemium business model and the introduction of new tools have made analytics pervasive in the video game industry. The research on game analytics is scant and descriptive. Thus, reasons for employing game analytics are not well understood. In this study, we analyze data collected with a set of in-depth interviews from small and medium-sized freemium game developers. The results show that game analytics is used to (1) assist design, (2) to reduce the risks associated with launching new games, and (3) to communicate with investors and publishers. The study advances the research on the business value of game analytics.

Keywords: Analytics · Freemium · Games

1 Introduction

Use of big data and analytics has become pervasive in the video game industry [1]. The adoption of analytics has been driven by the fast development of cost-effective solutions that enable basic analytics even for start-up sized game developers. The second driver of game analytics is the diffusion of the freemium business model that has created a need to accurately measure, predict, and intervene player behavior[1]. For example, Supercell, the company behind the top-grossing freemium games Clash of Clans and Boom Beach, generated revenue of 1.8 billion USD from in-game purchasing in 2014[2].

The term freemium is a combination of words "free" and "premium" [2] that describes a business model in which a service or a product is offered a for free but a premium is charged for advanced features, functionality or related products and services. Game analytics refers to applying analytics and big data in the gaming context [3]. Game analytics can be used to improve the players' gaming experience, or to maximize in-game purchases by tweaking various aspects of the game [4, 5]. For example, the time taken for a specific task, the cost of a specific item or the power of a specific weapon [6, 7] can be optimized with game analytics.

[1] http://www.theguardian.com/technology/2014/dec/09/clash-of-clans-billion-dollar-mobile-games.

[2] http://www.theguardian.com/technology/appsblog/2013/jul/23/clash-of-clans-supercell-freemium.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 326–337, 2015.
DOI: 10.1007/978-3-319-25013-7_26

Game analytics is a new field with limited research coverage [8]. Prior research has focused on describing methods of data gathering and analyzing [3] as well as the role of analytics in game development [9, 10]. The limited amount of prior literature can be partly explained with the fact that game developers consider analytics confidential and are thus often reluctant to share information on their analytics processes [11].

As a result, the business aspects of game analytics are not well understood. Consequently, this study seeks to investigate *why and how do small and medium-sized freemium game developers use game analytics?* To answer the research question we have conducted a set of in-depth interviews among freemium game developers. Our results show that game analytics is used to (1) assist design, (2) to reduce the risks associated with launching new games, and (3) to communicate with investors and publishers. The study advances the research on the business value of game analytics.

2 Literature Review

The Freemium Model. The freemium business model is extensively employed with digital products or services such as software, games, and web services. As a result, companies employing the freemium model strive to monetize their user base by striving to convert the users of the free version into paying customers [12–14].

Freemium games typically employ micro-transactions and advertising as the main monetization strategies [15]. Micro-transactions refer to buying virtual items or services that can be used and have value only inside a specific gaming environment [cf. 13]. Second, the players can be provided with the opportunity to buy "time", either to bypass waiting enforced by the game mechanics or to skip repetitive "grinding" phases [15], or to unlock additional levels or areas of the game. Third, virtual items and benefits compared to the non-paying users can be bundled into a premium user account or a subscription [12]. Fourth, in-game advertising and potentially offering an ad-free upgraded version against a fee can be used to monetize the players. Fifth and finally, game operators can sell information about the players to third parties, typically marketers [16].

Freemium games are typically designed to engage the player immediately, because there is no initial cost causing a lock-in effect [17]. In addition, the game mechanics used in freemium games motivate the players to make in-game purchases [18]. As a result, sustained play and customer lifetime value are critical for the economic success of freemium games.

Game Analytics. Game analytics is a subset of analytics applied to the game development [19]. Analytics in turn refers to using business intelligence in the process of discovering and communication patterns from the data and using the recognized patterns in solving business problems [20].

The need for game analytics has increased as games have become more sophisticated and complex [21], and the rise of mobile gaming and the freemium business model has also had its effect. Gross et al. [22] maintain that studying online games is important from a managerial viewpoint "to understand ways that interactions while gaming can be improved, in order to make better games."

Today's video games, particularly online games and social media games, can collect data about almost all players' in-game activities [4]. Game developers can use this data to obtain information about e.g. potential sources of player frustration [23]. For example, Replica Island, a game for Android devices, employed a player tracking system to identify instances where players were facing difficulties (e.g. player deaths). The whole metrics system was implemented by a one-man team at virtually no monetary cost, and has proven to be extremely valuable in identifying problems with the game design [23]. However, according to Drachen et al. [24], game analytics is far from the traditional value chain of the video game development and hence not very highly prioritized.

Prior game analytics research has employed the purchase funnel concept to illustrate the challenges the freemium model poses [3, 15, 25]. The AIDA-model (awareness, interest, desire, and action) from consumer behavior literature describes the process of new product adoption [26]. In application marketplaces where people typically download freemium games the customer can go through the process from awareness to action in seconds. However, out of the people who have downloaded a freemium game, only a fraction, e.g. 5 per cent will pay anything [15]. As a result, successful employment of the freemium model requires sustained player engagement and efficiently managing a large pool of non-paying players towards conversion during the course of the play. A freemium game can be a hit without ever being profitable, and thus monetization is essential [27].

3 Empirical Research

Data Collection. The empirical data was collected with five semi-structured interviews from experienced professionals from companies developing freemium games. The companies they worked in were based Finland. Obtaining the empirical data was challenging since finding the number of seasoned professionals with in-depth knowledge on freemium games as business in general and game analytics in particular is small. At the time of the interview, each of the five interviewees was in the process of developing a freemium game and using software to collect and analyze gameplay data. In addition, topics related to game analytics were frequently considered company confidential and hence many prospective informants declined the interview.

The theme of the interview was presented to the interviewees in advance. A semi-structured approach was selected due to the relative novelty of the research area. According to May [28] the semi-structured approach allows the interviews to be flexible and interactive.

The interviews were done face-to-face in sessions from fifty minutes to one hour twenty minutes. All interviews were recorded and transcribed by the first author. In addition, notes were taken during the interviews. Appendix 1 summarizes the data collection.

Analysis. The analysis process started already during the interviews as notes about potentially insightful themes were taken. In the first round of analysis, the interviews were transcribed and coded based on the research question [28]. Through iterations of the data, a constant-comparative method was applied to identify, elaborate, and clarify

categories [29]. Emergent categories were examined within and across interviews to determine salience and recurrence. Interconnections and discrepancies between the interviews and previous literature and between different interviews were also coded. The recurrence of certain key themes and limited number of new ones emerging over coding process indicated that the data exhibited saturation.

Results. We identified two main themes that describe the different roles of game analytics for small and medium-sized freemium game developers. These are analytics as a communication tool and analytics as a decision support tool. Based on these two themes we derived an emergent theme, analytics as necessity.

Analytics as Commnunication Tool

> "[Analytics] are a kind of tool for studios to justify their decisions, for example why a certain game is not ready for launch yet, because we need to improve this metric. The investors are more willing to give extra time, when they can see that in the long run the game will make more money if improved." (P4)

As the quotation above illustrates, investors and publishers are very keen on analytics. The interviewees clearly stated that the key performance indicators have to be reported frequently to investors and publishers, and lack of improvement in these metrics will lead to questions. On the other hand, analytics help game designers make their case when they feel the launch schedule needs to be postponed as the game is promising but behind schedule.

> "If I were a game publisher, I would ask teams to soft launch and provide retention and ARPDAU [average revenue per daily average users] numbers before I would invest anything." (P3)

Freemium games are typically continuously improved through scheduled updates and developers want to see the impact of changes as soon as possible. Sizeable and potentially risky adjustments are tested out with a small subgroup of players so that regular service would not be disturbed. Similarly, before the game is launched worldwide, it can be released in smaller market (e.g. Canada or Finland). The purpose of these soft launches is to collect data to ensure that the game will be profitable.

> "We used to have a thing where once a week every studio [under that publisher] would report to the headquarters in California in an hour long conference call and give a pre-formatted presentation in which the key metrics were analyzed, future plans to improve them laid out etc." (P4)

The direct business benefits of game analytics include informed financial decisions, such as rationalizing marketing spending and budgeting for launch. The interviewees were very conscious that the video game industry has had several high-profile costly flops, i.e. projects that went overtime and over budget and made considerable losses. The interviews also indicated some of which could potentially have been avoided with the help of analytics. In addition, the interviewees recognized pressure from many stakeholders, competitors, players, and publishers on the gaming companies to adopt game analytics.

> "Retention is the most important metric in the game industry" (P2)

The interviewees considered retention rate as the most important individual metric of business success in freemium game development. Consequently, all interviewees maintained that they are keen on tracking and monitoring retention figures. Retention rate refers to how players keep playing a game for subsequent sessions. A simple example of measuring retention is tracking the number of game overs per player. Typically retention was measured over a time period (e.g. 7 days, a month). The interviews also revealed that retention is a key metric for monetization as well as the basis of funnel analysis. In addition, retention is used to improve first impressions and the tutorials.

Interestingly, yet there was a consensus regarding importance and benefits from game analytics, the interviews also indicated that increasing volumes of data make it more challenging to extract relevant information from it. The interviewees wanted to keep the analytics process as simple as possible and focus on the key performance indicators (KPIs) and their trends over time and different versions.

> "The most common wisdom in this free-to-play model is that if you don't have retention you are never going to make money. Retention stems exactly from that the game itself has some interesting aspects and is in some way fun."(P4)

In addition to retention, metrics that measure monetization and additional retention metrics, e.g. average revenue per user (ARPU), conversion rates, tutorial funnels and day 1, day 7, day 14 and day 30 retention rates, were mentioned as examples of metrics that most game developers monitor, or at least should monitor.

Additionally, customer lifetime value is used to measure to what extent the costs related to customer acquisition are covered. Customer lifetime value is an aggregate of other metrics, namely cost-per-install and ARPU, and is used to guide marketing spending. The following quotation depicts the role of customer lifetime value:

> "It is mostly based on things like is it profitable to invest in marketing the game. LTV (Lifetime Value) will tell you that." (P1)

Analytics as a Decision Support Tool

> "You cannot make a good game with just analytics, it's very challenging – [everything new] comes from the creative side" (P2)

The informants, particularly the ones who had been actively involved in actual game design, stressed quite strongly that analytics does not, and it also should not, drive game design. The interviewees maintained that analytics can support decision-making by e.g. occasionally disqualifying the intuition of the designers, but that analytics does not offer ready solutions. Further underlining that the design philosophy is not data-driven but data-supported, the interviewees stressed that game developers can and will outsource data collection, crunching and storage, but never game design. The analytics and metrics do not make the games, but they can help in making them better. The following quotation describes the role of analytics in relation to design:

> "[Regarding Game Analytics] maybe the larger benefit, at least in our case, is that instead of driving development, they are used to spot errors in the code and clear design mistakes. - - It's more about monitoring – first you design, then you code and then you monitor how well did it go." (P5)

We observed an interesting controversy regarding the interviewees' views on the role of game analytics. While all believed in the value of analytics and some emphasized tracking as much as possible to ensure having sufficient amount of data whenever needed, others preferred a more strict selection of metrics to avoid information overload and collecting what they referred as "vanity metrics". These interviewees also more deliberately emphasized the importance of evaluating the benefits against the respective costs:

A recurring theme in the interviews was that the informants considered players unpredictable and that they often behave differently than the designers expected, often seemingly irrationally. Analytics can help designers to understand players' perspective of the gaming experience. For example, one interviewee described a mobile game that had a design mechanic punishing players for erroneous behavior by reducing his/her points. An alternative solution was to end the game immediately after an error was made and force the players to start over. Somewhat counter-intuitively the latter version, which the designers felt was more "hardcore-oriented", was more appealing to players and led to higher retention levels. As a result, game analytics proved that, against game conventions and designers' expectations, the harder and more punishing version was more popular.

The analysis tools and methods in use were largely uniform among informants. Maybe for this reason they did not see analytics as a real source of competitive advantage. As all our informants worked for small or medium-sized companies with very limited human and financial resources to be spend on analytics, they had simply adopted 'off-the-shelf' analytics tools. Since most small and medium-sized game companies use a similar set of tools, the informants considered the expertise and proficiency in using tools and the ability to draw the right creative conclusions the best way to derive value from game analytics. The informants also stated that the use of analytics differs considerably between smaller and larger companies in the freemium games market.

There was a strong consensus among the informants that data allows real and accurate insights about player behavior. All of the interviewees had experience from primarily quantitative analysis and metrics. The informants' view was that game companies seldom utilize qualitative data, as it was considered more taxing, less effective, and harder to implement with third-party solutions. The informants also stated that interviewing players about their gameplay habits can lead to misleading results since people can seldom tell what they really want and would use.

> "We don't do analytics because they are cheaper, but because they are better. – – I never trust people who say "If you would develop this, I would use it all the time". Only when I can measure that they really use it, I will believe it." (P3)

All informants stated that the analytics data is actively shared and communicated within the development team, but the interviews also indicated that game companies seldom share their sales numbers or metric data in public. Companies can however compare the KPIs of new launches with their prior games. This way game analytics also help guide the portfolio management of game publishers.

The data-supported design process is iterative. First, a visible problem is noticed in high-level metrics. Then the designers seek for possible causes by drilling down into the data. Specific changes are made to the game, and then the effects are measured. There is a constant loop for validating design decisions.

> "[We tend to find a] – –high-level issue and then try to find one specific user experience issue that you think you change, and then you iterate the process." (P4)

A typical way to tackle the issues that are found during the development process is to create two or more different versions of the game. These versions are then randomly distributed to players and their respective performance is measured. This is referred to as A/B testing. It cannot be used extensively for every decision since developing each alternative consumes resources and the inferior versions are a wasted effort. Acquiring relatively reliable results from the A/B tests also takes considerable time.

> "When something new is added to the game, it is done in two different ways and half of the players get version A and the other half gets version B and then the metrics are compared. There may be a hypothesis behind the test, but it is more about trying to find what works and then developing the better version further, leave it as is or abandon both." (P1)

Funnel analysis was another common analysis method used by the informants. Tutorials are a typical example of a funnel. Throughout the tutorial phase gradually fewer and fewer players reach each subsequent step. They have not necessarily paid anything for the game yet, so it is essential that they do not churn out this early. Measuring the return rate or retention of players in subsequent steps (or levels) in the game was the most common way of analyzing the quality of said funnels. When the designers note a notch in the retention, they know that there is a problem in that part of the funnel.

Analytics as a Necessity. The findings strongly suggest that game analytics plays an important role in communicating with investors and publishers as well as in supporting game development and manager's decision-making. Hence, analytics was seen as a necessity in freemium game development. Every informant pointed out several benefits of game analytics. In fact, operating without utilizing analytics was considered "flying blind" and analytics was seen as a means to reduce the risk of failures. Interestingly, the interviewees also repeatedly stated that there had been no need to state an explicit business case to justify investments in game analytics – the benefits were that clear.

> "It is clear that – especially in this free-to-play model – you cannot operate under the mentality that you just launch a game and hope for the best." (P1)

However, utilization of game analytics is characterized by the lack of resources. This explains why game analytics was seen as something that is important but not as a differentiating factor or a potential source of competitive advantage. Table 1 summarizes the main findings.

Table 1. Summary of findings

Analytics as a communication tool	Investors and publishers follow certain key metrics The metrics provide a common ground for discussion Retention is the most important metric to track Average revenue per user (ARPU) and Lifetime Value (LTV) are also actively followed
Analytics as a decision support tool	Help to reduce the risk of total failure Role is to support design decisions, not drive them Analytics affect the whole development process, there is a constant loop of changes, assessment and improvement Developers use game analytics to e.g. identify problems in the game design and/or bugs in the code that lead to player churn
Analytics as a necessity	Analytics viewed necessary but not as a source for competitive advantage Reliance on third-party game analytics tools and software. Lack of resources (time, skills, and money) restricts the use of more advanced analytics The emphasis is on quantitative data

4 Discussion

Theoretical Implications. The findings indicate that game analytics play a pivotal role in freemium game development. The interviewees emphasized a combination of designer creativity and analytics is required to obtain the best results and repeatable results. Our results imply that retention is the most important metric to track. Game analytics was not viewed as source of competitive advantage, at least for smaller companies relying on standard third-party analytics tools, but considered more as a risk reduction tool. The possible competitive advantage of utilizing game analytics is derived from the experience and creativity of the analysts and designers.

Prior research has presented the vast possibilities and introduced sophisticated solutions for game analytics [1, 3]. However, our results imply that the analytics processes used by small and medium-sized freemium developers are rather simple. There are two opposing philosophies in data gathering. One stresses that data cannot be tracked retroactively, and tracking as much as possible ensures you have the data when you need it. The other warns about collecting "vanity metrics" that do not contribute to game design and only confuse the analysis process [see e.g. 30].

Very large volumes of data call for increasingly sophisticated and powerful analysis tools, which in turn increase the costs. Prior research has promoted utilizing a combination of data sources (e.g. analytics, interviews, biometric measuring) to improve results [10]. However, our results show that small and medium-sized freemium developers primarily focus on measuring retention and a few other key metrics. Furthermore, the data analysis methods were simple and many of the professionals cited that they do not have advanced database management skills.

Altogether, these observations align with prior research arguing that alongside being actionable, the results from analytics should be human-readable and easily interpretable [as in e.g. the model presented in 31].

Retention rates were further utilized in evaluating which alternative was better in A/B tests and improving funnels, especially tutorials. The methods used resembled the split testing used by Andersen et al. [32] to recognize the most engaging design choice from three alternatives. The small and medium-sized freemium developers also showed interest in adapting more complex analytical procedures, but did not have either the skills or resources to implement them. The interviewees did not share much about the gathering of qualitative data, which further emphasizes that the focus is on directly actionable, quantitative metrics.

Prior literature on game analytics has recognized the managerial importance of actionable results [33]. Our results add on this body of research by emphasizing the value of the predictability and decision support in analytics. In addition, our results demonstrate that for freemium game developers, game analytics provided important tools for investor communications. Third, game analytics can help in optimizing marketing spending by customer lifetime value calculations or campaigning for more time from the publisher citing promising retention trends.

Managerial Implications. Compared to game industry's prior reluctance to adopt analytics [7], our results indicate a change in mindset. Our set of interviews among game industry professionals indicated that the benefits from game analytics are so self-evident that they are not even always explicitly stated. Since basic analytics are today available for even startup companies, the initial costs of implementing game analytics are relatively low. For example, applying basic telemetry data analysis is one the most cost-efficient way to do user research, and even one-man teams can afford it [34].

Our results imply that game analytics is viewed as a risk reduction tool. Using analytics during the development stage can decrease the risk of total failure in the launch stage. In addition, even the basic level of analytics with the standard key performance indicators can assist companies in e.g. terminating projects that are unlikely to generate sufficient revenue to become profitable.

Certain metrics have become close to industry standards. Key performance indicators that measure monetization and retention, such as ARPU, conversion rates, tutorial funnels and day 1, day 7, day 14 and day 30 retention and customer lifetime value were mentioned as examples of the key metrics. These metrics are also used to evaluate and predict the financial success of the game.

Many of the interviewees actually stated that investors require key ratios such as retention rate and ARPU when they discuss potential investments with the teams. Similarly, publishers expect results and want hard numbers to confirm the feasibility of the

game concept. For example, certain leading publishers also already demand weekly reports that cite key metrics from their developers.

Limitations and Future Research. Like any other empirical research, the present study is subject to a number of limitations. First, the empirical data was collected solely from small and medium sized companies developing freemium games that were based in Finland. For example, large freemium developers such as Supercell are well known for their advanced use of game analytics. Thus, we suggest future research with a broader empirical coverage.

Second, the freemium model as well as analytics is employed in other fields such as online music as well as more traditional fields such as insurance. Future research could examine how analytics is utilized in other sectors to find best practices that could be applicable across industries.

Appendix 1: Table of Interview Subjects

	Position	Description	In development	Company size
P1	Product Lead	Experience from multiple mobile game companies. Has been utilizing game analytics since 2009.	Freemium mobile game	SME 20-49 employees
P2	Chief Product Officer	Has led his own game studio for many years. First touched simple forms of analytics in 2006, has been responsible for analytics at many companies.	Freemium mobile game	SME 5-9 employees
P3	Chief Executive Officer	Started a gaming company after an engineering career. Begun with analytics in 2011 and has utilized them in two successful game projects.	Freemium mobile game	SME 5-9 employees
P4	Product Manager	Begun as a community manager, and utilized analytics in that role. Has now shifted focus to e.g. balancing in-game economy.	Freemium mobile game	SME 20-49 employees
P5	Chief Executive Officer	Started a small game development company with friends. Few launched games; the new project is the first time analytics are used systematically.	Freemium PC game	SME 5-9 employees

References

1. Bauckhage, C., Kersting, K., Sifa, R., et al.: How players lose interest in playing a game: an empirical study based on distributions of total playing times, pp. 139–146 (2012)
2. Wilson, F.: The Freemium business model. A VC Blog, vol. 23 (March 2006)
3. El-Nasr, M.S., Drachen, A., Canossa, A.: Game Analytics: Maximizing the Value of Player Data. Springer Science & Business Media, Berlin (2013)
4. Drachen, A., Thurau, C., Togelius, J., et al.: Game data mining. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game Analytics, pp. 205–253. Springer, Berlin (2013)
5. McAllister, G., Mirza-Babaei, P., Avent, J.: Improving gameplay with game metrics and player metrics. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game Analytics, pp. 621–638. Springer, Berlin (2013)
6. Zoeller, G.: Game development telemetry in production. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game Analytics, pp. 111–135. Springer, Berlin (2013)
7. Mellon, L.: Applying Metrics Driven Development to MMO Costs and Risks. Versant Corporation, Redhood City (2009)
8. Drachen, A., Canossa, A.: Towards gameplay analysis via gameplay metrics, pp. 202–209 (2009)
9. Canossa, A., El-Nasr, M.S., Drachen, A.: Benefits of game analytics: stakeholders, contexts and domains. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) game analytics, pp. 41–52. Springer, Berlin (2013)
10. Gómez-Maureira, M.A., Westerlaken, M., Janssen, D.P., et al.: Improving level design through game user research: a comparison of methodologies. Entertainment Comput. **5**, 463–473 (2014)
11. Wallner, G., Kriglstein, S., Gnadlinger, F., et al.: Game user telemetry in practice: a case study, vol. 45 (2014)
12. Mäntymäki, M., Salo, J.: Why do teens spend real money in virtual worlds? A consumption values and developmental psychology perspective on virtual consumption. Int. J. Inf. Manag. **35**, 124–134 (2015)
13. Mäntymäki, M., Salo, J.: Purchasing behavior in social virtual worlds: an examination of Habbo Hotel. Int. J. Inf. Manag. **33**, 282–290 (2013)
14. Wagner, T.M., Benlian, A., Hess, T.: The advertising effect of free–do free basic versions promote premium versions within the freemium business model of music services?, pp. 2928–2937 (2013)
15. Fields, T., Cotton, B.: Mobile & Social Game Design: Monetization Methods and Mechanics. Morgan Kaufmann, Los Altos (2012)
16. Clemons, E.K.: Business models for monetizing internet applications and web sites: experience, theory, and predictions. J. Manag. Inf. Syst. **26**, 15–41 (2009)
17. Zauberman, G.: The Intertemporal dynamics of consumer lock-in. J. Consum. Res. **30**, 405–419 (2003)
18. Hamari, J., Lehdonvirta, V.: Game design as marketing: how game mechanics create demand for virtual goods. Int. J. Bus. Sci. Appl. Manag. **5**, 14–29 (2010)
19. Drachen, A., El-Nasr, M.S., Canossa, A.: Game analytics–the basics. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game analytics, pp. 13–40. Springer, Berlin (2013)
20. Davenport, T.H., Harris, J.G.: Competing on Analytics: The New Science of Winning. Harvard Business Press, Cambridge (2007)
21. Hullett, K., Nagappan, N., Schuh, E., et al.: Data analytics for game development: NIER track, pp. 940–943 (2011)

22. Gross, S., Hakken, D., True, N.: Getting real about games: using ethnography to give direction to big data (2013)
23. Pruett, C.: Hot failure: tuning gameplay with simple player metrics. Game Dev. Mag. **19** (2010)
24. Drachen, A., Canossa, A., Sørensen, J.R.M.: Gameplay metrics in game user research: examples from the trenches. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game Analytics, pp. 285–319. Springer, Berlin (2013)
25. Moreira, Á.V., Vicente Filho, V., Ramalho, G.L.: Understanding mobile game success: a study of features related to acquisition, retention and monetization. SBC **5**, 2–12 (2014)
26. Webster, F.E.: New product adoption in industrial markets: a framework for analysis. J. Mark. **33**, 35–39 (1969)
27. Davidovici-Nora, M.: Paid and free digital business models innovations in the video game industry. Commun. Strateg. **94**, 83–102 (2014)
28. May, T.: Social Research: Issues, Methods and Process. McGraw-Hill Education, UK (2011)
29. Corbin, J., Strauss, A.: Strategies for Qualitative Data Analysis: Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory. SAGE Publications, Thousand Oaks (2008)
30. Canossa, A.: Meaning in gameplay: filtering variables, defining metrics, extracting features and creating models for gameplay analysis. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game Analytics, pp. 255–283. Springer, Berlin (2013)
31. Xie, H., Kudenko, D., Devlin, S., Cowling, P.: Predicting player disengagement in online games. In: Cazenave, T., Winands, M.H., Björnsson, Y. (eds.) CGW 2014. CCIS, vol. 504, pp. 133–149. Springer, Heidelberg (2014)
32. Andersen, E., Liu, Y., Snider, R., et al.: On the harmfulness of secondary game objectives, pp. 30–37 (2011)
33. Canossa, A., Cheong, Y.: Between intention and improvisation: limits of gameplay metrics analysis and phenomenological debugging. In: DiGRA Think, Design, Play (2011)
34. Canossa, A.: Interview with Nicklas "Nifflas" Nygren. In: El-Nasr, M.S., Drachen, A., Canossa, A. (eds.) Game Analytics, pp. 471–473. Springer, Berlin (2013)

Dynamic IT Values and Relationships: A Sociomaterial Perspective

Leon Dohmen(✉)

CGI Nederland B.V., Eindhoven, Netherlands
Leon.dohmen@cgi.com

Abstract. Management scholars are criticized for ignorance and the wrong approach when studying the impact of technology in organizational life. Impact of technology in this paper is interpreted as IT values created or achieved from equivalent and contingent interaction between human (people) and non-human agents (technology, organization). Researchers and theorists propose to include a sociomaterial perspective and to develop general and broader, empirical based patterns across different contexts. Based on a literature review containing publications of theoretical considerations and empirical research this paper introduces a first general and sociomaterial based overview and taxonomy of IT values and their relations. IT values have a techno-economic or socio-techno orientation, are dynamically entangled and competitive, and complementary or overlapping. IT values are related to time, sponsor and, hierarchy. The identified IT values are ordered into a framework which has to be treated as a starting point to discuss further the definition, dynamics and relations of IT values from a sociomaterial perspective.

Keywords: Impact of technology · IT values · Sociomateriality · Relationship · Entanglement · Emergence · Techno-economic · Socio-techno

1 Introduction

Management scholars are criticized in two ways when researching the *impact of technology* in organizational life. Either they ignore it[1] or they prefer a linear approach by separating technology, organization and people [1]. Sociomaterial theory is proposed as alternative. "*Sociomateriality* stands out as a symbol for the interest in the social and the technical, and in particular, the subtleties of their contingent intertwining" [2]. One of the key concepts of sociomateriality is based on Actor Network Theory. Human and non-human agents are inseparable connected maintaining equivalent relationships and "enact continuously relational effects" [2]. But also sociomateriality is subject of criticism. Sociomaterial oriented theorists and researchers argue to unlock broader, general patterns across different contexts [2] and to acquire more empirical evidence.

[1] Orlikowski refers to a study of Zammuto et al. including a survey of four journals: Academy of Management Journal, Academy of Management Review, Administrative Science Quarterly and, Organization Science. Only 2.8 % of the research articles in these journals focused on technology and organizations.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 338–353, 2015.
DOI: 10.1007/978-3-319-25013-7_27

This paper interprets impact of technology as *IT values* achieved from the equivalent and contingent interaction between technology, organization and people. To meet the objection of linear research, IT values will be analyzed through a sociomaterial lens. First IT values will be identified, collected and ordered based on literature review and discussions with subject matter experts. Secondly this paper introduces the relational dimensions between collected IT values.

2 Dynamic IT Values through a Sociomaterial Lens

Assumed is that IT values are depending from the sociomaterial context and are emergent as attitude, learning processes and skills play a crucial role [3]. Here we accept that values are in a "state of becoming" rather than a status quo [2] assuming a fully relational ontology, where IT values exist in relation to each other.

To acquire insights in the dynamics of IT values when applying IT-facilities this study, identifying and collecting IT values, proposes a new sociomaterial based IT values framework and taxonomy. The question we focus on for this inventory is: *Which IT values can be identified and how are they related?*

2.1 Research Method

To search for broader and general sociomaterial patterns zooming out technique is used for a literature review containing theoretical expositions and/or empirical research. A zooming approach provides ideas about how to extend qualitative research methods for investigations of sociomateriality [2]. Exploration and selection of IT values and composition of the framework and taxonomy happened from March 2011 until April 2015. The study started with a consultation of subject matter experts with an academic and/or business background. Following their literature suggestions revealed a heterogeneous representation of the concept of IT value.[2] Due to the zooming out approach publications from outside the IT domain like product design were included. Books and articles showed different approaches and definitions to describe and explain IT value (related) concepts. These concepts showed for example quantitative and qualitative research approaches and/or objective and subjective (perceived) definitions of IT value. Another difference was that some articles focus more on the process of value creation and conditions rather than on IT value. Because the contribution of this paper is to compose a framework and taxonomy for IT values we focused especially on studies and considerations with primary attention for explaining and defining IT value. After composing the first draft of the framework further consultation of subject matter experts took place which led to new literature suggestions. Additional literature review was done using keywords (see footnote 2) searching on the Internet for relevant articles,

[2] Recommended literature, which some of is included in the list of references, led to an extensive list of concepts related to or representing IT value: business/IT alignment, business value, company value, contribution of IT, customer satisfaction, customer value, information system success, IT absorption, IT adoption, IT value perception, net benefits, technology acceptance, user acceptance, use of IT, user value and, value of IT.

papers and books dealing with the subject of IT value. New IT values discovered became part of the collection. To construct and order collected IT values, terms and definitions of the public values discussion [4] are adopted for the IT values framework. Public values are values in governance and public service [4]. The public values discussion involves the relational and entangled dimension between values which corresponds with important principles of sociomateriality. Strong point of this long term and evolving zooming out approach is the broad coverage of the concept of IT value. A weak point is due to the extensive range of related key words (see footnote 2), important relevant literature can be missed. Despite this weakness, because of the need to unlock broader general patterns as contribution to support sociomaterial theory it is chosen for this approach.

The framework presented in this paper consists of eight nodal values, sixteen neighbour values and sixty four co-values (Table 1).[3] Neighbour values are the bridge between nodal values, which are referred to as starting point and co-values are determined as promoter or contributor [4]. Co-values can be positive or negative. The interpretation of an IT value and position in the framework is partly derived from the categories of user value from Boztepe [5].[4] Because of the readability of the table references are omitted here. These are explicitly mentioned in the accompanying text.

We prosecute our discourse about IT values with a brief description and definition of chosen values when appropriate supported by illustrative examples from practice.

2.2 Utility Value

Utility value is the consequence of using a product and encompasses neighbour values *convenience*, *safety*, *quality (and performance)*, *economy* [5] and, *service* [6]. IT values to a large extend are connected to the material aspects of the product and have a techno-economic orientation.

A utility is acquired to fulfil (convenience) needs – including *accessibility*, *appropriateness* and (physical) *compatibility* – of the user and *avoid unpleasantness* [5]. Technological *availability* [7] should also be considered as an important co-value for convenience. The human role is shaping as well as being shaped by *time* [8]. Orlikowski and Yates [8] define time as "people produce and reproduce what can be seen to be temporal structures to guide, orient and coordinate their ongoing activities."

Security, *health* and *comfort* are co-values of safe usage [7]. Paro, the robot seal used in healthcare for people suffering from dementia, is an interesting reference in practice for these co-values [9]. Paro reduces stress and leads to a positive mood. Serving road safety, *compliance* is meeting governmental laws and regulations [10].

Durability and *reliability* of materials used are important co-values of quality and performance [5]. *Efficiency* is also performance related [5]. Investments in basic

[3] Selected values follow the definition of the referred to literature in the reference list, in some cases supported by an example or key word between brackets.

[4] Boztepe originally distinguishes four categories which are positioned as nodal value in this paper: utility, social significance, emotional and spiritual. These categories are used as starting point for the framework and further supplemented with other relevant nodal values.

Table 1. General sociomaterial IT values framework

Nodal IT values	Neighbour IT values	IT co-values
Utility	Convenience	– Accessibility – Appropriateness – (Physical) compatibility – Availability – Time management – (Avoidance of) sensory unpleasantness
	Safety	– Security – Health (e.g. reducing stress) – Comfort – Compliance
	Quality and performance	– Durability – Reliability – Fit for purpose (usefulness) – Agility (flexibility) – Speed – Effectiveness – Efficiency (ease of use)
	Economy	– Use economy – Purchase economy (price value) – Objective financial indicators (e.g. net margin, profitability, operational expenses, etcetera) – Share value
	Service	– Assurance – Responsiveness – Empathy – Relationship
Social	Social prestige	– Influence – Power – Impression management (face saving acts) – Respect
	Identity	– Role fulfilling – Group belongingness
	Ethics	– Right conduct – Moral principles – Honesty
Emotional	Pleasure	– Fun – Enjoyment – Beauty – (Job)satisfaction – Attachment – Affection (love) – Detachment – Addiction – Nomophobia

(*Continued*)

Table 1. (*Continued*)

Nodal IT values	Neighbour IT values	IT co-values
		– Panic – Anger
	Sentimentality	– Memorability – Nostalgia
Cognitive	Stimulation	– Excitement – Curiosity – Self-actualization
	Growth	– Independent thought and action (independence) – Creating new innovative things – Diffusion (of gained knowledge)
Universal	Welfare	– Social innovation – Tolerance
	Protection	– Sustainability – Care for people
Traditional	Loyalty	– Commitment (deep attachment) – Respect
Spiritual	–	– Good luck – Superstition
Singularity	Super-humanity	– Super-intelligence – Immortality – Personalized food

infrastructures have a different purpose compared to investments in innovative applications [11]. So the utility should fit the *purpose* of use. Business operation changes permanently due to increased competition, new rules in law, etcetera. To adapt smoothly to these changes *agility* [12] or *flexibility* is another important aspect of quality and performance [13]. Increasing *speed* to access knowledge and service delivery is experienced as an important gain when applying social tools [14]. Al-Maskari and Sanderson [15] refer to a general term like *effectiveness* to express utility value. From an expectancy perspective Venkatesh et al. [16] refer to ease of use (*effort*) and *usefulness* (performance). Usefulness is interpreted as similar meaning as fit for purpose.

Economy is a next neighbour value of utility value used in as well as the business context [9, 17, 18] as consumer context [5, 16]. Where Boztepe [5] refers to *purchase economy* Venkatesh et al. [16] introduce the term *price value* as an indicator of technology use in a consumer context. Sneller [9, 18] and Kersten [17] emphasize the importance of *use* economy when discussing utility value. Kersten [17] urges to replace legacy systems ('old' IT) by modern technologies ('new' IT). Economy value includes the life cycle of a technology and is measured with *objective financial indicators* like *net margin* and *profitability*. Implementing an ERP-system increases *share value* [18].

Besides applying a utility or product, service is determined as an important neighbour value of utility value. DeLone and McLean [6] adapted their previous IS

Success Model. Service quality contains similar co-values as identified for product quality and performance. Additional important co-values are *responsiveness*, *assurance* and *empathy*. *Relationship* between actors when providing services to a product also impacts user value perceptions [19].

2.3 Social Value

Social value involves socially oriented benefits. From a sociomaterial perspective here the material is used to derive or gain social advantage. This includes *social prestige* and construction and maintenance of one's *identity* [5]. *Ethics*, another identified neighbour value, is of increasing importance [20].

"Social significance (prestige) value refers to the socially oriented benefits attained through ownership and experience" [5]. Product benefit examples lead to social associations (*impression management*, *face saving acts*) between family and other social groups with increase in *respect*, *influence* and *power* as consequence [21]. "Possession of a trendy object is often seen as sufficient to communicate a certain image of self (identity)" [5]. Social significance or influence becomes meaning in relations with others which concerns *group belongingness* and *role fulfilling* [5, 21]. Companies can build up a company image (identity) by chasing IT fashions [22].

Ethics refers to a set of (local) principles of *right conduct* or a theory or system of *moral* values. IT solutions have a big impact on the work of others. Engineers of IT solutions should embed an ethical (value) dimension (*honesty*) in the requirements of the to build solution [20].

2.4 Emotional Value

Emotional value is about aroused feelings of affective states like *pleasure* and is triggered by co-values like *fun* and sensory *enjoyment* [16, 21]. Also hedonistic values like *beauty* initiate pleasurable experiences and belong to this neighbour value pleasure. *Memorability* can arouse a *sentimental* feeling which is also associated with emotional value [5].

Attachment is a positive (pleasurable) emotional state in the relation between user and product [23]. Opposite to it *detachment* is a negative emotion which indicates the lack of linkage between an individual and a product [23]. Socio-technical studies see *job satisfaction* and productivity as important outcomes manipulated by social and technical factors [7]. Automation leads to controlling and deskilling while empowering and upskilling are a result of informate. Both – automate and informate – are different purposes when applying IT leading to different values [7]. Robot seal Paro is an interesting example how people suffering from dementia can get *attached* to it and develop *affection* for Paro [9]. On the other hand attachment can evolve into habit for example in mobile phone usage [16] and *addiction*. Venkatesh, Thong and Xu [16] introduce habit for technology use in a consumer context to extend UTAUT.[5] Habit

[5] UTAUT is Unified Theory of Acceptance and Use of Technology.

here is defined as prior behaviour or automatic behaviour. *Nomophobia* [24] is detected as a new 'illness'. When people have lost or forgotten their mobile phone emotional feelings like *panic* become part of them.

Using a screensaver or background picture on a device with a family photo or *memorable* event – *nostalgia* – is an experience of emotional feelings. Sentimentality is here appointed as a separate neighbour value [5].

Emotional value benefits arise from affective experiences related to aesthetic, giving meaning and provoke feelings as *love* (affection) and *anger*. Emotional co-values are person related, subjective and intrinsic [5]. These contributors are mainly assigned to the socio side of sociomateriality.

2.5 Cognitive Value

"IT has given a boost to knowledge related activities which are a continuation of the written word and printed book. This has been sometimes referred to as the information revolution" [25]. Heng [25] provides a classification scheme for IT-applications and addresses the *cognitive* value of IT. This nodal value is lacking in the overview of Boztepe [5]. The value of IT recognized as a source to contribute to knowledge creation and distribution is linked to the network era [26], social technologies and the networked organization [27]. *Stimulation* and *growth* are neighbour values of cognitive value. Cognitive value is also designated as epistemic value [21].

"IT's contribution (stimulation) to the knowledge enterprise enable employees to create, store and disseminate knowledge on a scale hitherto unknown" [25]. *Curiosity, excitement* and *self-actualization* are important elements to acquire (new) knowledge and support the *creation* and *growth* of *new (innovative) things* [21]. Value sensitive design here is linked *to independent thought and action* decoupled from group values [21]. The more rapidly (individual) innovative IT capabilities can be deployed, the more rapidly (business) value will grow [28]. While legacy research highlights random adoption in a social network Baldwin and Curley [28] claim that *diffusion* of innovations can be actively directed and accelerated, especially for IT systems.

Cognitive value – diffusion and access to information and knowledge – is a value which evolved from efficiency and information value [26]. Cognitive value is primarily intrinsic and subjective but can be made explicit and objective. Applying IT in education becomes more and more popular and impacts the learning process [29]. Savas [23] appoints *independence* as a positive emotion possibly leading to attachment to a product. This link to emotional value shows that values are not perceived isolated but are dynamic and (closely) interwoven [5].

2.6 Universal Value

To *care for people* in emerging countries platforms on the Internet like Get It Done [30] provide opportunities to fund projects or create your own projects. These crowd funding initiatives supported by IT are occurrences associated with *universal* values like *welfare* and *protection*. Welfare and protection can also be associated with public values [4]. However, the discussion about public values goes far beyond IT values

only. The application of IT by national and local administrations leads besides IT value also to transformations in relationships between governmental institutions reciprocally and to transformations in relationships between governmental institutions and their citizens [31]. The latter is another example of dynamic emergence [2] of IT values.

Social innovation [32] and *tolerance* reflect welfare. People are able to connect with social tools [27] to whoever they like to. The application of IT is also seen as a social innovation issue in Belgium to support the increasing aging of the baby boom generation [33].

Sustainability shows care for planet and nature. A great example of protection is the 'volmeld'-system of the city Groningen in the Netherlands [9]. Underground waste containers are equipped with a system which transfers twice a day a message of the degree of filling to the central computers. If the percentage reaches seventy percent the container is automatically included in the route of the garbage trucks. Besides time savings (efficiency) which is usually a business objective the environmental pollution is decreased due to the reduction of co2 emission by seventeen percent [9]. U-city concerns the environmentally friendly and sustainable smart (or knowledge) city which makes the ubiquitous computing available amongst the urban elements such as people, building, infrastructure and open space [34]. This is an ongoing example of the evolving entanglement of technology, things and people like described and explained in sociomateriality theory and the dynamic values it creates [2].

Universal value is associated with care for people and planet [21]. From a socio-materiality perspective this nodal value is very much linked to the socio part of materiality. However, these values can easily be mixed with organizational value objectives like improving efficiency.

2.7 Traditional Value

Respect and *commitment* are related to acceptance of customs and ideas that *traditional* culture and religion impose on themselves [21]. Commitment (deep attachment) is related to co-value attachment which is associated with nodal value emotional.

Besides involving content also the support in users' tasks in maintaining ideas and customs is an example of traditional value. Commitment can evolve to *loyalty* to a product and recommendation. Commitment can also lead to repeat purchase of a product [35] or increased (intention of) use [16]. "Positive experience with use will lead to greater user satisfaction in a causal sense. Similarly, increased user satisfaction will lead to increased intention of use and thus use" [6].

Within traditional value the socio and material are very close related. Irritation and frustration – as experiences of negative value perceptions – can be linked directly to the (material) product when it does not work properly or as expected.

2.8 Spiritual Value

"*Spiritual* value refers to *good luck* and *sacredness* (*superstition*) enabled by a product" [5]. According to Boztepe [5] examples show that communication technologies are increasingly becoming enablers of spiritual experiences too. For instance, several

websites have been set up that serve Muslims who live away from their home countries, allowing them to pay online and make sacrifices on their behalf.

2.9 Super-Humanity Value

Technological *singularity* [36] which can be achieved via biomedical science and nanotechnology will create *super-intelligence* and *super-humanity* [37]. Although mainly envisioned and to date hardly proven, the concept seems to be able to abolish biological limitations and create immortality [38]. Vinge [37] described several appearances of singularity and super-human being based on artificial intelligence, intelligence amplification and, biomedical, Internet and, digital Gaia[6] scenarios. For this paper, singularity is appointed as a provisional end point in the evolution of IT (see Fig. 1).

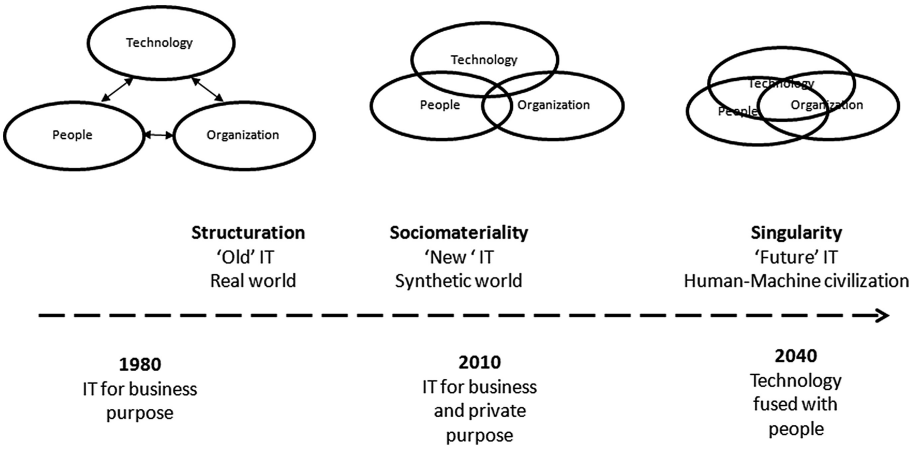


Fig. 1. Emergent IT values and time relation

A further application of IT that should not be ignored is 3D food printing [39]. Data-driven recipes (Jeffrey Lipton, Cornell Creative Machines Lab) provide possibilities for older people with chewing and swallowing problems to create a meal with a modified food structure[7] and nutritional hardness (Pieter Debrauwer, TNO). 3D (food) printing adds a new dimension to the application of IT. While old applications and IT systems transfer data into reports and documents – focusing mainly on data, transaction and transfer – 3D printing transforms data into physical products. Hereby 3D food printing creates also new sensory experiences. The printed food can be tasted and smelled. Other values mentioned during the 3D Food Printing Conferences [39] were

[6] The Digital Gaia Scenario: The network of embedded microprocessors becomes sufficiently effective to be considered a superhuman being.

[7] In the food industry this is also referred to as food texture.

contribution to flexibility, sustainability and comfort. Instead of feeding elderly people with pureed food which sometimes leads to social isolation while people feel ashamed, feeding people with *personalized food* that looks regular but which structure is adapted increases group belongingness (nodal value social). To date it can be determined that IT values are able to trigger all five human sensory experiences: feel; see; hear; smell, and; taste.

With this involvement of biological (food) aspects it could be considered to include the biological aspect in sociomaterial theory. Also the nodal value universal and neighbour value protection is linked to biological value when involving a co-value like sustainability.

2.10 The Sum of IT Values: Net Benefits

With respect to the ongoing theoretical and philosophical discussion of foundations for sociomateriality [40] our study to explore, define and categorize IT values contains a broad orientation and generalization leading to an overview and ordering of IT values from a sociomaterial perspective. Herewith we respond to the call and need for sociomaterial generalization across different contexts [2].

Same as for public values [4] IT values seem to be in competition and are not strictly separated but overlapping each other. From utility value – being the material representative – a number of socio related IT values are triggered, sometimes as an undesired and negative side-effect like addiction or when improving efficiency consequence can be a decrease in working pleasure – informate versus automate [7]. This means that IT values do not exist in isolation, are dynamically entangled, maintain *relationships* and are the sum up of positive and negative values in a certain context. To express the sum of positive and negative IT values we propose to connect the term *net benefits* [6, 41] to sociomaterial theory when discussing and researching IT values. Net benefits have to be defined depending on the context, are probably the most accurate descriptor of value and, should address the level of analysis [6].

The term cyborg[8] [2] is suggested as a useful metaphor that supports exploration and explanation of sociomaterial reality in future organizational life. But we also observe a further entanglement or even fusion of the synthetic[9] and real world [1]. IT values are in an ongoing state of becoming subjected to dynamics of emergence [2].

3 IT Values and Relationships

As explained before values do not appear in isolation but are closely interwoven [5] and subject to dynamics of emergence [2, 3]. In our interpretation of the relational ontology *relations* between values can be competitive, complementary or overlapping

[8] A hybrid of machine and organism. See also: http://www.zdnet.be/nieuws/161067/belgische-cyborg-opent-deuren-met-ingeplante-chips/.

[9] Orlikowski [1] here refers to the world known as MKP20 within Sun Microsystems. Sun Microsystems is acquired by Oracle in 2010. See also: http://virtualworldsforum.com/.

[4]. The terms nodal, neighbour and co-values [4] suggest that IT values are related and entangled in a value network. At least three types of relationship can be distinguished in our relational ontology: time, stakeholder and hierarchy. We zoom in further to these relation types to discover, describe and explain the relational tension between IT values.[10]

3.1 IT Values and Time Relation

The value of IT shifts over time from focus on efficiency to focus on effectiveness and flexibility and customer satisfaction [26, 42]. Explaining the difficulties and complexity of IT an article in the Dutch professional magazine CFO refers to 'old' and 'new' IT [17]. Examples mentioned for old IT are enterprise resource planning and mainframes. Social media is referred to as a new IT trend. Pictured by Fig. 1, let us put the emergence of IT and IT values in a *time* perspective. We include the theoretical origin and expected continuation of sociomateriality.

The start of this timeframe is around 1970 denoted as the fifth technological revolution [43] and a (provisional) destination is called singularity whose estimated range is around 2040 [37, 38]. In the beginning of the technological revolution IT was applied mainly inside organizations only for business purposes. IT from that era like enterprise resource planning applications, mainframe systems and terminals is typed as old IT. Around 1990 the structuration theory – theory of structure and agents – supports social research focusing on the impact of IT on business processes and work executed by people within organizations [7]. Much has changed in the decades after 1980.

Sociomateriality saw light as successor of structuration theory because technology, organization, and people became more entangled and inseparable [44]. IT became widespread into as well as organizational as private lives. Besides physical presence IT offers the possibility for virtual presence which introduced the phenomena of synthetic world [1]. In synthetic worlds, deployed within organizations, people can collaborate and communicate real time [1]. IT created in this era is referred to as new IT in this paper. Appearances of new IT are social media and mobile devices like smart phone and tablet computer. This is also the era the paper positions phenomena like digital business, services and society: IT for business and private purpose.

But the end is not reached yet. Researchers and scientists predict that people and technology will fuse which will create new and unprecedented opportunities which are referred to as 'future' IT. IT converged with other technologies will overcome biological and cognitive limitations and interaction between humans and non-humans will change dramatically [36–38].

IT was once a mean to improve operational efficiency, to date IT has changed in a multipurpose and multiform vehicle covering a broad spectrum of socio and material related values. The entanglement and dynamic emergence of IT and IT values clarify the objection to linear and quantitative research [1]. Referring to the general IT values

[10] This paper contains a first orientation, interpretation and description of the sociomaterial relation between IT values. Further theoretical and empirical research is needed to determine this sociomaterial relational concept in more detail.

framework in Table 1 we can conclude that the framework contains a mix of old (e.g. efficiency), new (e.g. social innovation) and future (e.g. super-humanity) IT values.

3.2 IT Values and Stakeholder Relation

Besides time related IT values are also assigned to different *stakeholders* [45]. Reviewing and re-specifying the DeLone and McLean model of IS success the term net benefits [41] distinguishes different types of stakeholders: societal, organization (management and groups) and individual. IT values listed in Table 1 depend from the stakeholder point of view and can vary by situation. Stakeholder relation is an important source for competition between values. When striving for organizational value as efficiency this can lead to deskilling [7]. Involving IT outsourcing causes multiple customer and supplier stakeholder relations.

Chau et al. [45] compared IT value studies conducted in Asia and Europe. They note that in IT value research most studies involve the organizational level. Fewer studies focus on the individual level. Chau et al. [45] distinguish objective measures – e.g. accounting and financial indicators, costs, return on investment and, firm value – besides perceptual measures – e.g. increased decision quality, better alignment with business strategy, etcetera. They observe a general shift from using objective measures to perceptual measures to study IT value.

A special entry to societal value is the debate about public values [4]. This includes at one hand the broader discussion about the common good referring to contributors to value like public interest and social cohesion. On the other hand national and local governments apply technology serving public values. Technology shapes the intra-organizational aspects of public administration institutions and the relationship between public administration institutions and citizens leading to IT induced public sector transformation [31]. Due to increase of digital fraud and other crime (cyber-crime) security seems to become a value of increasing societal importance.

3.3 IT Values and Hierarchy Relation

The third relational dimension is *hierarchy*. Hierarchy is related to the relative primacy or importance of a value [4]. Relative primacy depends on the context. From a public value point of view liberty may be more important than efficiency. Within an organization, especially when acting in private competition, efficiency may be of more importance instead of employee's job satisfaction. Hierarchy of values and their relations are considered to be inseparable. Hierarchy between values is designated as prime values and instrumental values. Prime values are seen as temporary[11] conditions whereas instrumental values are consequences [4]. The hierarchical relation between IT values listed in Table 1 in this paper can be different per situation.[12]

[11] We add here the word temporary before condition because this reflects the sociomaterial principle that IT values are in an ongoing state of becoming subjected to dynamics of emergence.

[12] This relational view and entanglement of IT values – based on sociomaterial principles – is important and meaningful to understand the concept of IT values in this paper. However, more investigation is needed to deepen out this relationship and to create a stronger theoretical fundament.

Referring once again to Paro – the robot seal used in healthcare to accompany patients suffering from dementia – how should value be expressed? Paro is reducing stress and has a positive impact on the mood of patients. Should we primarily look at the aspects of value or should we first look at the cost of development and maintenance of Paro [9]? In the Paro example the techno-economic view [43] and socio-techno perspective [44] become unified.

4 Summary and Looking Forward

Primary contribution of this paper is to unlock general and broader patterns regarding the impact of technology which is interpreted as IT value. Herewith this paper encompasses a response to the call and need to fundament sociomaterial theory [2] because the concept of sociomateriality is extremely theoretical and philosophical [40]. A second purpose for contribution is to close a bit of the gap in management research ignoring the impact of technology in organizational life [1]. The study is limited to IT value as a result of the interaction between technology, organization and people. The process of value creation and conditions is excluded.

The answer to the research question "*Which IT values can be identified and how are they related?"* delivers a taxonomy and framework of IT values containing eight nodal, sixteen neighbour and sixty four co-values. The sociomaterial relationship between IT values is dimensioned to time, stakeholder and hierarchy. Grounding theories of sociomateriality like Actor Network Theory are part of ongoing discussions [2]. Due to dynamics of emergence and the continuous state of becoming the concept of IT values is a spectacular and interesting sociomaterial subject for social research. IT values, their entanglement and relations are far more than a linear relationship between two or a (limited) number of variables and their causality and therefore a difficult to capture phenomenon.

IT values are the result of interaction between technology, organization and people [1]. The result of this extensive study is a generalized IT value framework based on sociomaterial theory and guiding principles. This overview should be seen as a starting point for further discussing IT values and their relationships. The principles below are accompanying the general sociomaterial IT value framework (see Table 1):

1. The study to explore, collect, define and categorize IT values to a general framework is the answer to the call to search for generalizations and broader patterns to support sociomaterial theory.
2. The constructed framework should be seen rather as a starting point for further discussion and research than as an end point.
3. IT values have a socio-techno or techno-economic orientation.
4. IT values are measured subjectively (perceived) or objectively.
5. IT values are classified as nodal, neighbour and co-values.
6. Due to dynamics of emergence IT values should be understood as a state of becoming (temporary condition) ontology instead of a solid state.
7. IT values exist not isolated but are entangled and maintain dynamic emergent, competitive, complementary and overlapping relationships.

8. IT values are time, stakeholder and hierarchy related.
9. IT values can be either positive or negative.
10. Net benefits are the sum of positive and negative IT values.
11. It should be considered to include biological aspects into sociomaterial theory.

In a next paper the theoretical foundations of sociomaterial IT values and their relations are applied in a digital business situation. Subjects like new IT, mobile IT, user behaviour, use patterns, security and IT values in a digital culture will be deepened. This next paper will contribute to sociomaterial theory by associating empirical evidence to this paper dealing with IT values and their dynamics and relationships.[13]

Acknowledgements. The author thanks Bert Kersten, professor at Nyenrode Business University and Marijn Janssen, professor at Technical University Delft for their sublime and extended guidance and support during this study. Without this great help this paper could not have been written.

References

1. Orlikowski, W.J.: The sociomateriality of organisational life: considering technology in management research. Camb. J. Econ. **34**(1), 125–141 (2009)
2. Cecez-Kecmanovic, D., Galliers, R.D., Henfridsson, O., Newell, S., Vidgen, R.: The sociomateriality of information systems: current status, future directions. MIS Q. **38**(3), 809–830 (2014)
3. Deursen van, A.J., Dijk van, J.A.: CTRL ALT DELETE: Productiviteitsverlies door ICT-problemen en ontoereikende digitale vaardigheden op het werk. Enschede: Universiteit Twente/Center for e-Government studies (2012)
4. Jørgensen, T.B., Bozeman, B.: Public values: an inventory. Adm. Soc. **39**(3), 354–381 (2007)
5. Boztepe, S.: User value: competing theories and models. Int. J. Des. **1**(2), 55–63 (2007)
6. DeLone, W.H., McLean, E.R.: Information systems success revisited. In: Proceedings of the 35th Hawaii International Conference on System Sciences 2002, Waikoloa, Hawaii, pp. 2966–2976. IEEE (2002)
7. Orlikowski, W.J.: The duality of technology: rethinking the concept of technology in organizations. Organ. Sci. **3**(3), 398–427 (1992)
8. Orlikowski, W.J., Yates, J.: It's about time: temporal structuring in organizations. Organ. Sci. **13**(6), 684–700 (2002)
9. Sneller, L.: Over de waarde van IT. Nyenrode Business Universiteit, Breukelen (2012)
10. Janssen, M.: Technologische krachten in het bestuur. Technical University Delft, Delft (2013)
11. Goldstein, P., Katz, R.N., Olson, M.: Understanding the value of IT. EDUCAUSE Q. **26**(3), 14–18 (2003)

[13] Personal reflections, findings and insights collected during the study to IT values and their relations are captured and expressed in a number of (smaller) articles and blogs on media like Slideshare, ManagementSite and Blogit: http://www.slideshare.net/ldohmen; https://www.managementsite.nl/auteurs/leon-dohmen; http://www.blogit.nl/author/leon-dohmen.

12. Verniers, H., Teunissen, W.: IT Service Portfoliomanagement: Maximaliseer de Waarde van IT. Van Haren Publishing, Zaltbommel (2011)
13. Symons, C.: Measuring the Business Value of IT. Forrester Research, Cambridge (2006)
14. Bughin, J., Chui, M., Manyika, J.: Clouds, big data, and smart assets: ten tech-enabled business trends to watch. McKinsey Q. **56**, 75–86 (2010)
15. Al-Maskari, A., Sanderson, M.: A review of factors influencing user satisfaction in information retrieval. J. Am. Soc. Inf. Sci. Technol. **61**(5), 859–868 (2010)
16. Venkatesh, V., Thong, J.Y., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**(1), 157–178 (2012)
17. Kersten, B.: De werkelijke bijdrage van IT aan de business, pp. 20–23. CFO (2011)
18. Sneller, L.: Does ERP Add Company Value. HAVEKA, Alblasserdam (2010)
19. Sun, Y., Fang, Y., Lim, K.H., Straub, D.: User satisfaction with information technology service delivery: a social capital perspective. Inf. Syst. Res. **23**(4), 1195–1211 (2012)
20. van den Hoven, J.: Over duimschroeven en kant-en-klaarmaaltijden. de IT-auditor **3**, 5–7 (2011)
21. Kujala, S., Väänänen-Vainio-Mattila, K.: Value of information systems and products: understanding the user's perspective and values. J. Inf. Technol. Theor. Appl. **9**(4), Article 4 (2009)
22. Wang, P.: Chasing the hottest IT: effects of information technology fashion on organizations. MIS Q. **34**(1), 63–85 (2010)
23. Savas, Ö.: A perspective on the person-product relationship: attachment en detachment. In: McDonagh, D., Hekkert, P., van Erp, J., Gyi, D. (eds.) Design and Emotion, pp. 317–321. Taylor & Francis, London (2004)
24. Szpakow, A., Stryzhak, A., Prokopowicz, W.: Evaluation of threat of mobile phone - addition among Belarusian university students. Prog. Health Sci. **1**(2), 96–101 (2011)
25. Heng, M.: Three Dimensions of Information Technology Applications: A Historical Perspective. Vrije Universiteit, Amsterdam (1993)
26. van der Zee, H.: Business Transformation and IT. Dutch University Press, Tilburg (2001)
27. Bughin, J., Byers, A.H., Chui, M.: How social technologies are extending the organization. McKinsey Q. **20**, 1–10 (2011)
28. Baldwin, E., Curley, M.: Managing IT Innovation for Business Value. Intel Press, Hillsboro (2007)
29. Puentedura, R.R.: SAMR and TPCK: a hands-on approach to classroom practice. Ruben R. Puentedura's weblog (2014). http://www.hippasus.com/rrpweblog/archives/2014/12/11/SAMRandTPCK_HandsOnApproachClassroomPractice.pdf. Accessed 3 January 2015
30. Verboom, H.: Get It Done. Retreived on July 1, 2012, Get It Done: http://www.getitdone.org/ (2009)
31. van Veenstra, A.F.: IT-induced Public Sector Transformation. BOXpress, 's-Hertogenbosch (2012)
32. van den Broek, C., Dohmen, L., van der Hooft, B.: Changing IT in Six. Royal Van Gorcum, Assen (2010)
33. Mousaid, S., Leys, M.: ICT in Wel en Wee. Vrije Universiteit Brussel, Brussel (2012)
34. Lee, S.H., Han, J.H., Leem, Y.T., Yigitcanlar, T.: Towards ubiquitous city: concept, planning, and experiences in the Republic of Korea. In: Yigitcanlar, T., Velibeyoglu, K., Baum, S. (eds.) Knowledge-based Urban Development: Planning and Applications in the Information Era, pp. 148–170. Information Science Reference, Hershey (2008)
35. Lam, S.Y., Shankar, V., Erramilli, M.K., Murthy, B.: Customer value, satisfaction, loyalty, and switching costs: an illustration from a business-to-business service context. J. Acad. Mark. Sci. **32**(3), 293–311 (2004)

36. Vinge, V.: The Coming Technological Singularity. Feedbooks, Paris (1993)
37. Vinge, V.: Signs of the singularity: hints of the singularity's approach can be found in the arguments of its critics. IEEE Spectrum's SPECIAL REPORT: THE SINGULARITY (2008)
38. Kurzweil, R.: Reinventing humanity: the future of machine-human intelligence. Futurist **40**, 39–46 (2006)
39. 3D Food Printing Conference: The future of 3D food printing for professionals and consumers. In: 3D Food Printing Conference (2015). http://3dfoodprintingconference.com/. Accessed 20 April 2015
40. Leonardi, P.M.: Theoretical foundations for the study of sociomateriality. Inf. Organ. **23**(2), 59–76 (2013)
41. Seddon, P.B.: A Respecification and extension of the DeLone and McLean model of IS success. Inf. Syst. Res. **8**(3), 240–253 (1997)
42. Nelson, M.R.: Assessing and communicating the value of IT. Res. Bull. EDUCAUSE Cent. Appl. Res. **2005**(16) (2005)
43. Perez, C.: Technological Revolutions and Techno-Economic Paradigms. The other canon foundation, Norway and Tallinn University of Technology (2009)
44. Orlikowski, W.J., Scott, S.V.: Sociomateriality: Challenging the Separation of Technology, Work and Organization. London School of Economics and Political Science, London (2008)
45. Chau, P.Y., Kuan, K.K., Liang, T.-P.: Research on IT value: what we have done in Asia and Europe. Eur. J. Inf. Syst. **16**, 196–201 (2007)

Designing Viable Multi-sided Data Platforms: The Case of Context-Aware Mobile Travel Applications

Mark de Reuver[1(✉)], Timber Haaker[1,2], Fatemeh Nikayin[1], and Ruud Kosman[2]

[1] Delft University of Technology, Engineering Systems and Services, Delft, The Netherlands
{g.a.dereuver, t.haaker, f.a.nikayin}@tudelft.nl
[2] InnoValor B.V., Enschede, The Netherlands
{timber.haaker, ruud.kosman}@innovalor.nl

Abstract. Advances in data semantification and natural language querying are enabling new generations of context-aware mobile applications. Such applications would rely on platforms that integrate heterogeneous sets of user data from a range of applications and systems. Designing these platforms is challenging as they should serve multiple user groups at the same time. In this paper, we analyze who should subsidize multi-sided data platforms that enable mobile context-aware travel applications. After analyzing the different user groups and revenue models, we assess end-user acceptance of these revenue models through a survey among 197 potential users. Results show that users willing to share data with app developers are more inclined to use data-driven mobile travel apps but are less inclined to pay for them. This paradoxical result explains why premium-pricing as well as data-monetization strategies can both be viable.

1 Introduction

While app developers have long struggled to seize the opportunity [1], they now commonly use context information to enrich their mobile applications. While GPS is the most prominent examples, app developers also rely on aggregated past behavior, inferred preferences and past transactions. However, if app developers share such context information among each other, even smarter context-aware applications can be envisioned. New technologies for linking and semantifying large sets of data as well as natural language querying approaches are driving these opportunities. One promising application area is that of travel applications: whereas most existing traveling apps focus either on booking, transport advice or local events, sharing user context data between such applications would enable far more integrated and coherent travel advice.

Sharing context information generated by a mobile application with other app developers risks violating user privacy as well as harming competitive position of the individual app developer. These issues could be circumvented by instantiating shared and trusted data platforms that aggregate user data. However, designing such platforms is challenging since they should satisfy multiple user groups: end-users, app developers

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 354–365, 2015.
DOI: 10.1007/978-3-319-25013-7_28

and potentially even others. A core issue is then who should subsidize the platform, i.e. what is the source of revenues going to be [2, 3, 4]. User willingness to pay for more advanced context-aware travel advice is often problematic. Monetizing user data by selling it to third parties can be a risky approach as well.

The objective of this paper is *to analyze who should subsidize multi-sided data platforms that enable mobile context-aware travel applications*. We do so by analyzing the case of 3cixty: a data platform for mobile context-aware travel applications, developed within the context of 3cixty. We outline different revenue models and subsequently analyze user acceptance of such revenue models by analyzing the results of a survey among 197 early adopters in the Netherlands.

Section 2 develops a theoretical background based on multi-sided platform literature. Section 3 provides a background on the 3cixty case and technical platform. Section 4 sketches potential value networks and revenue models, which are subsequently assessed through user research in Sect. 5. Section 6 concludes the paper, draws limitations and implications for future research.

2 Background on Multi-sided Platforms

In general, platforms can be seen "as building blocks (they can be product, technologies or services) that act as a foundation upon which an array of firms can develop complementary products, technologies or services" [5]. A platform serving multiple user groups, such as buyers and sellers, is typically denoted as a multi-sided platform [6]. The objective of a multi-sided platform is to facilitate the transactions between different user groups, such as consumers and app developers. Multi-sided platforms can be analyzed using the framework of two-sided markets [3]. The focal artifact in this paper can be seen as a multi-sided platform as it connects multiple groups of actors (i.e., end-users and app developers) while providing generic functionality on which services can be developed (i.e., shared databases and querying tools).

Such shared platforms can generate considerable network externalities, which implies that the value of the platform depends on the number of users [7, 8]. Network externalities imply that the value of a system depends on the installed base of users [9, 10]. Increasing adoption levels can thus lead to a positive feedback cycle that further increases the usefulness of the technology [11].

Network externalities are direct if the value of the product increase by others buying, connecting, or using the same platform or services provided via the platform. The utility of the platform increases with the number of other using it. Examples of direct network effects are social media, which become more valuable if more end-users join the platform. But also because platforms have interchangeable components, users can share the benefits of the same technical advance. Backward compatibility, interoperability and interface standards are therefore crucial. Typically, direct network effect refers to positive effect between users in a group; however, when more consumers for a platform reduce the value for similar consumers, the platform entails a negative direct network effect.

Externalities are indirect when the value of the platforms depends on the number of users in a different user group. For instance, video game consoles become more

valuable for consumers if there are more developers creating games for that console. Indirect network effects may also be negative, for instance more advertisers on a search engine platform decrease its value for searchers of independent advice. These indirect effects are typical for multi sided markets, where service providers, developers and consumers meet. As service providers as well as developers make the platform's various components better, the platform gets more attractive over time. As consumers use the platform more, they make the market larger. Once the adoption of a product or technology has started, these network externalities provide benefits to both new and existing users such as reduced price, lower uncertainty about future versions of platforms as well as complementary services, communities of users, higher quality products, new market opportunities.

Network externalities are important since they call into question which user group should subsidize the platform. If one user group of the platform is considered to be of more value than the other, it may be that they are subsidized by lowering prices or offer access for a fee [3]. In practice, different subsidization models are being used. The concept of marquee user specifically refers to those user groups that have such great value for the other user groups that their adoption of the platform should be subsidized [12].

3 The 3cixty Case

The objective of the project is to provide a data platform for apps and services for city visitors (i.e., tourists as well as citizens). The platform will retrieve data from websites (e.g. hotels, restaurants, events, sights), other platforms (e.g. social media), smartphones (e.g. apps for exploring a city) and/or sensors. The platform offers various data services to app developers, for instance a cross domain querying language, a crowdsourcing mechanism, a mobility profiling service, query augmentation and social media mining.

Apart from these enabling services, the platform provides a clean data repository about a city that can empower specific applications. The data is "semantified" (i.e. semantics of the data has been made explicit) and "reconciled" (i.e. heterogeneous data coming from various sources has already been integrated). App developers can access this clean data repository via a single application programming interface (API) which is web developer friendly (i.e. apps developer do not need to interact with many specific data sources APIs). App developers can access this clean data repository via queries which provide an additional level of expressivity and enable to provide answers to queries that no single data source can do. The platform can answer questions like "Give me the list of hotels reachable within 10 min by metro of the venue of the Franz Ferdinand concert". The specific features of the platform are listed in Table 1.

Table 1. 3cixty platform features

	Frequency of city trips
High-level querying language with a cross-domain knowledge base	Execute mixed-/cross-domain (tourism, mobility, etc.) queries in your app that combine diverse types of information, including the types of interlinked data provided by the other 3cixty services
Query augmentation service	Automatically personalize queries with restrictions based on data provided by other services (e.g., reviews by friends of users)
Generic crowdsourcing mechanism	Efficiently extend the 3cixty core APIs to enable users to contribute information about aspects of the city. Access such information provided by others and through on-demand crowdsourcing, ask people in real-time to contribute data to help others
A "parallel exploration" graphical user interface	Enable users to construct and save trees of interrelated queries that enable them to explore several aspects of the city simultaneously
A "wish list" service	Allow users to indicate where they may want to go and (optionally) when and how Store this information in the cloud so that it can be accessed on appropriate occasions, even from other devices
Social media mining	Enable users to see 'nearby buzz', i.e. what people are talking about on social media in the neighborhood or in the city
Mobility profiling service	Track users' movements within the city, including their use of modes of transportation Give users access to information from their mobility profiles - and, with permission, to those of relevant other persons

Using enabling services on the platform and the clean data repository, app developers can develop applications such as city trip and accommodation planning, traffic update, cultural and entertainment updates and information about the city.

4 Revenue Models

To roll out and commercialize the platform, different roles are required. The generic value network shows the business roles (blocks) and the value exchanged between them in the form of service and revenue flows (arrows). Typically different roles are required in the research phase and roll out phases of the services, see Fig. 1.

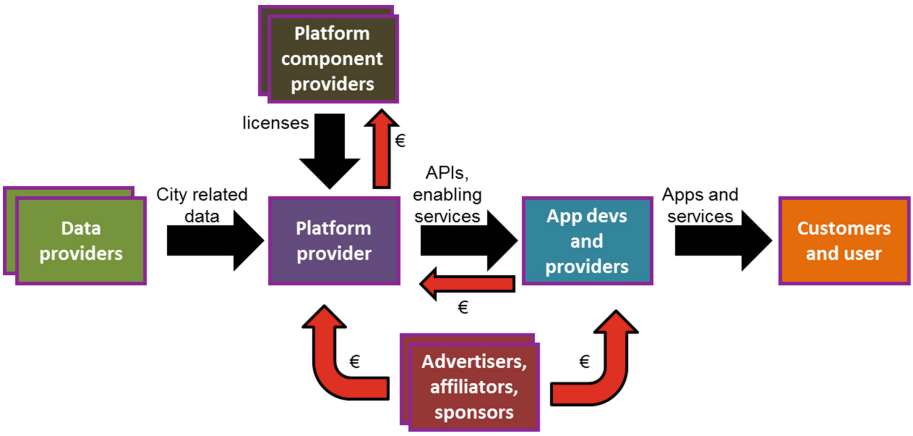


Fig. 1. Basic value network

Note that the 'user' is the actual person that uses the app while the 'customer' is the party actually paying for the app. For example a 'mobility solution provider' could act as service provider and commission an app developer to develop a 3cixty enabled mobility app. The app could be part of a mobility solution that is offered to companies ('customers'). The company's employees ('users') would use the app to manage their mobility. Of course the roles of service provider and customer need not be separated; some party may simply order an app with an app developer and offer it to users directly via an app store. Or an app developer may even directly develop and offer an app to users; in that case app developer, app owner/service provider and customer are the same actor. Further detailing of business roles is possible as well, as for instance the roles of app owners and service providers can be separated (accommodating for white label business models).

The production of the app is a joint effort of the parties on the left hand side of the picture with the platform provider (system integrator) playing a pivotal role. Several parties provide information to the platform which can, amongst other things, be stored, combined, enriched and aggregated in the platform. Possibly, the services could be subsidized or even paid for by finance providers like insures, (local) governments or other providers. As an extra revenue stream, advertisements could play a role in the value network as well. Advertisement agencies could offer profile based advertisements to their customers and the advertisement provider is in practice often a value chain in itself.

For defining the revenue model, two main issues have to be dealt with. First, it should be decided whether the platform is offered in a profit or non-profit approach. Second, it should be decided whether the platform is open or closed to third party app developers. The motivation for this distinction is that open and closed platforms as well as privately and publicly owned platforms occur in practice and provide relevant but distinct options for the 3cixty platform. In the closed model the platform is used internally as a service platform and in the open model the platform is offered 'as-a-service' to 3rd parties like app developers. The different scenarios are depicted in

Fig. 2 together with some actual examples of platforms and/or apps that follow such business scenario. One may also consider a situation in which there is no platform provider at all, but where (some of) the developed 3cixty software is made available as open source. In that case any interested organisations could use the software.

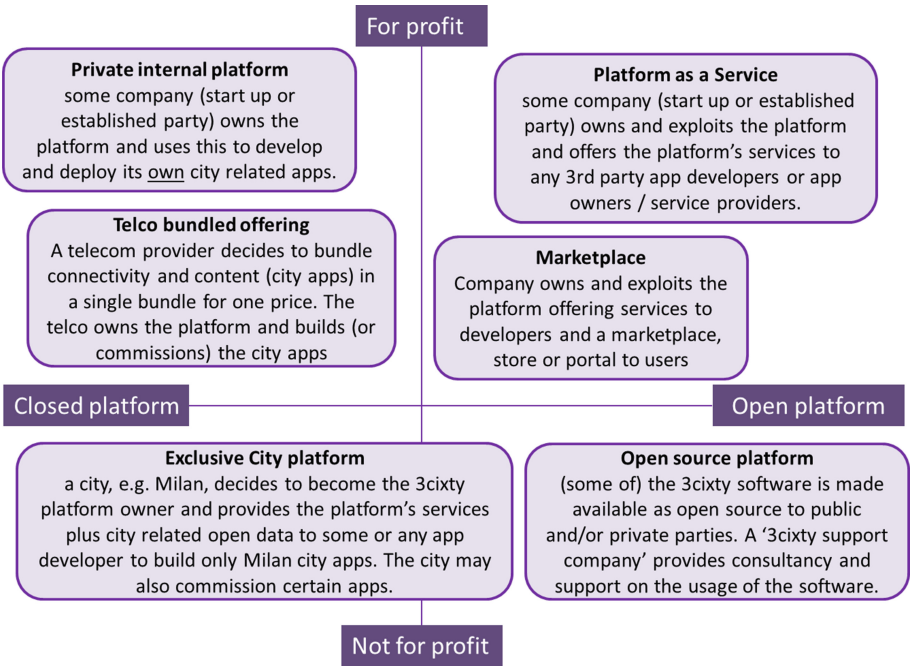


Fig. 2. Exploitation scenarios

Several revenue models can be considered to commercialize the platforms. The app market is dominated by the freemium model as over 90 % of all apps are for free, so user willingness to pay is often low. One approach to monetizing a service platform is that the platform is owned by some app developer and used as an internal platform to create and deploy its own apps. In this scenario the roles of platform provider and app developer are played by the same actor. The platform is closed in business sense, i.e. it can only be used by the app developer owning the platform. Other app developers or service providers cannot make use of the platform to build apps. Many providers use service platforms to efficiently develop and deploy their services and apps. A different approach is that the roles of platform provider and app developer are separated. In this case the services of the 3cixty platform are made available to app developers/service providers to support the creation and deployment of apps. Third parties can use part or all of the 3cixty functionality and data to develop new apps or enrich existing apps. Within these two scenarios still many business models are possible with different actors taking up the role of platform provider (private vs public party) and supported by

different revenue models, i.e. bundling, pay-per-use, affiliation, advertisement, subsidy and sponsoring, open source etc. However, the idea behind the development of the 3cixty platform is geared towards an open model.

Given these features many revenue models could be applicable. Affiliation models where commission fees from companies that appear in the traveling app and that realize transactions, e.g. bookings, via the app. Advertising models, in which contextual advertising based on user profiles (e.g. preferences following from searches or a wish list) and levels of context awareness (location, time). Or data driven model: local government, tourist board or tourist trade organisation may be interested in aggregated data from user preferences and behaviour. For any of the revenue models identified, user willingness to use 3cixty types of applications, willingness to pay for applications as well as willingness to share personal data with app developers and third parties are crucial issues.

5 Validation with Users

For any of the revenue models discussed in the previous section, a key assumption is willingness of end-users to use mobile travel apps that are based on integrated sets of context data. To further validate the revenue models, key concerns are user willingness to pay for context-aware travel applications as well as willingness to share data with platform providers. Such data could both be actively shared (e.g. crowdsourcing) or passively (e.g. past transaction information or location data). We test these assumptions through a survey among potential end-users for the platform and its applications.

5.1 Method

The intended population comprises young people who have a smartphone and who have a habit of traveling to cities. A convenience sample of Dutch students is used. Invitations to participate in the survey were posted on the Facebook page of various students of a bachelor course in December 2014. To obtain a homogenous group, we only include students possessing a smartphone in our sample. 197 valid responses were received, most of them being bachelor students. Age varies between 17 and 29, with average 20.6 years old. Gender was balanced with 53 % male. Respondents make 2.3 city trips per year, with standard deviation 2.1 and a maximum of 12.

Constructs were operationalized into self-developed, reflective scales, see Table 2. Although respondents were not exposed to the exact 3cixty service mockups, they were asked to reflect on 3cixty type of applications that provide comprehensive travel information and booking possibilities. All items were measured on a 7-point Likert scale. Confirmatory factor analysis was carried out using WarpPLS, see Table 1. Convergent validity is acceptable with average variance extracted exceeding .5 benchmark, and standardized factor loadings exceeding .6. Construct reliability is acceptable.

Table 2. Measures.

Construct	Measure	Std factor loading	AVE	Construct reliability
Frequency of city trips	I like to go to large cities on vacation	.83	.69	.82
	I like to go on vacation within Europe	.83		
Willingness to receive local travel experiences	I find earlier travel experiences of others at the same location relevant for my travel	.83	.68	.75
	When on vacation, I am open to suggestions about local activities	.83		
Willingness to actively share travel experiences	I like to explore the area when on vacation	.76	.58	.81
	When on vacation, I like to share my travel experiences	.76		
Willingness to passively share travel data	Please indicate to what extent you are willing to share with a mobile application: Location (e.g. for restaurants/sights nearby, routes)	.67	.50	.74
	Social media (Facebook, Twitter for a.o. opinions on sites, events etc.)	.80		
	Transaction history (for a.o. offers based on transactions, preferences etc.)	.64		
Intention to use	Assume a mobile app would be offered which combines existing travel apps (e.g. finding hotels, restaurants, public transport, events). Would you use such an application?			
Willingness to pay	How much would you be willing to pay for such an integrative travel app?			

Discriminant validity is acceptable as correlations between latent variables do not exceed the square root of average variance extracted, see Table 3.

5.2 Results

A structural model is assessed using WarpPLS. The advantage of using WarPLS is that it allows not only interval but also categorical and nominal variables, such as gender. The model shows acceptable overall fit according to the fit indices generally

Table 3. Correlations among latent variables (diagonals show square root of average variance extracted).

	Frequency of city trips	Willingness to receive local travel experiences	Willingness to actively share travel experiences	Willingness to passively share travel data
Frequency of city trips	.83			
Willingness to receive local travel experiences	.30	.83		
Willingness to actively share travel experiences	.34	.35	.76	
Willingness to passively share travel data	.07	.05	.06	.71

recommended with WarpPLS (Tenenhaus GoF = .359, Sympson's paradox ratio = .846), and low multicollinearity (Average block VIF = 1.215), see Fig. 3.

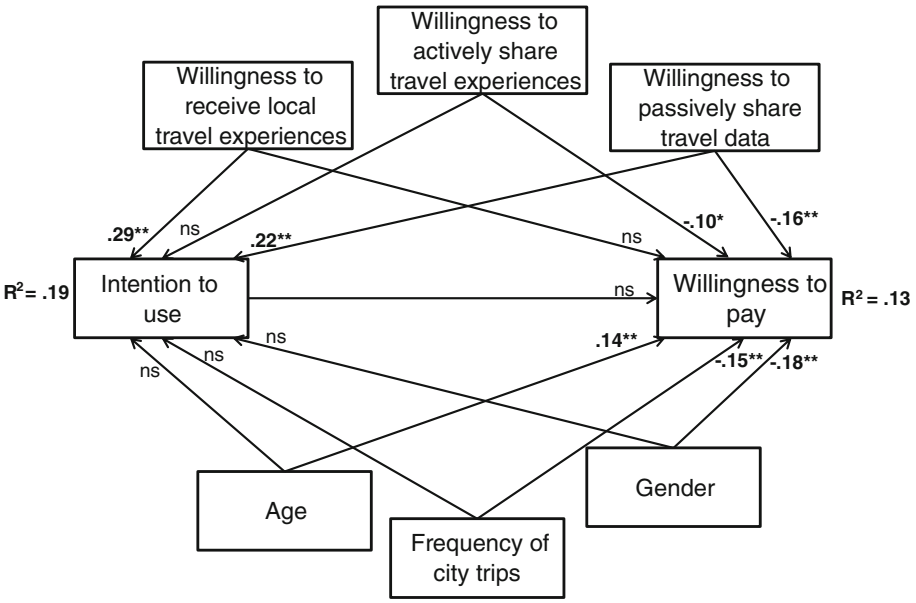


Fig. 3. Structural model (standardized regression weights on the arrows) * $p < .05$; ** $p < .01$

Explained variance of both intention to use and willingness to pay is low. We find that intention to use depends on the degree to which respondents plan their trips (in advance or ad hoc) and the degree to which they are inclined to share data via social media. On the other hand, willingness to pay depends on travel frequency, gender and age, while willingness to share data has a negative impact. Intention to use does not significantly affect willingness to pay.

The tendency to share data via social media also affects the intention to use. This implies that users of a travel application will be more likely to share data, which opens opportunities to utilize such data for monetization.

Willingness to pay, on the other hand, depends on how frequent a person travels, as well as gender and age. Usage intention does not contribute to willingness to pay, which suggests that even if people are willing to use the application, they are not willing to pay for it.

Interestingly, intention to share has a negative impact on willingness to pay. Apparently, those that are less inclined to share data on social media are more willing to pay for a travel app.

Another striking finding is that intention to use the travel app has no significant impact on willingness to pay for the application. Overall, the findings illustrate a dilemma for this type of social media traveling apps. Factors that drive intention to use completely differ from factors that drive intention to pay. This explains why alternative revenue models need to be considered in which app providers do not depend on willingness to pay. Apparently, one should either focus on those willing to pay for a travel app, but are not willing to actively share their data via social media. And when one focuses on people that are willing to share their data, willingness to pay will generally be lower. As such, the findings suggest a dilemma between two revenue models: a free app with monetization of shared data or a paid app without sharing of data with third parties.

6 Conclusions

Designing data platforms for context-aware mobile services entails several issues that are typical for multi-sided platforms. For the case being studied, we identified at least four different user groups that may or may not subsidize the platform: end-users, advertisers, app developers and government organizations. Who should subsidize the platform strongly depends on the exploitation scenario, i.e. whether a profit or non-profit model is chosen and whether the platform should be open or closed.

To evaluate the revenue models developed in this paper, we studied the end-user (i.e. consumer) perspective. We found support for a typical paradox observed in this domain: the reasons why consumers are willing to use applications are completely different from the reasons why they would be willing to pay. In fact, the more a user is willing to share data and travel experiences, the less that user is willing to pay for a context-aware travel app. As such, providers of data platforms can target one segment of end-users with a premium-priced app or another segment with a for-free app in which data is being shared for commercial purposes with third parties.

The findings are in line with earlier research on privacy concerns in information systems which suggests that people make trade-offs between utility, price and privacy harm of an application [13, 14]. How such compensation of privacy concerns with discounts or added value from apps works out in a big data era where consequences of disclosing personal data cannot always be foreseen warrants further research.

The paper contributes to understanding of the complexity in designing multi-sided data platforms. Platforms that link data providers, end-users, application developers and advertisers create highly complex sets of network externalities between and within the different user groups. In the case studied in this paper, it became clear that without prior demarcations on profit-orientation and openness, a multifold of revenue models can be considered in parallel. Even after our user research, different revenue models are still relevant.

We used a convenience sample gathered through links on social media. As our research model included willingness to share travel experiences including via social media, there may be systematic bias in the sample. However, considering the high adoption rate of social media in the population of young technical students, we assume this will not be a major threat.

The present paper focused on end-users for the validation of the revenue models. Our next step will be studying the app developers' perspective. Earlier workshop and informal interviews with app developers suggest they may be willing to pay for data platforms if they truly add value for their business. Whether this is the case depends strongly on technological issues such as how to structure the data, how to form the APIs and which of the technical features of the data platform become available first.

Acknowledgements. We are grateful for comments and suggestions from Anthony Jameson. The paper has been developed within the EIT Digital innovation activity 3cixty (14523). We appreciate the contributions by several BSc-students from the course TB232A at Delft University of Technology in collecting data for the user study reported in this paper.

References

1. De Reuver, M., Haaker, T.: Designing viable business models for context-aware mobile services. Telematics Inform. **26**, 240–248 (2009)
2. Armstrong, M.: Competition in two-sided markets. RAND J. Econ. **37**, 668–691 (2006)
3. Rochet, J.-C., Tirole, J.: Platform competition in two-sided markets. J. Eur. Econ. Assoc. **1**, 990–1029 (2003)
4. Rochet, J.-C., Tirole, J.: Two-sided markets: a progress report. RAND J. Econ. **37**, 645–667 (2006)
5. Gawer, A.: Platform dynamics and strategies: from products to services. Platforms, Markets and Innovation, pp. 45–76. Edward Elgar Publishing, Cheltenham (2009)
6. Evans, D.S.: Some empirical aspects of multi-sided platform industries. Rev. Netw. Econ. **2**, 19 (2003)
7. Church, J., Gandal, N.: Platform competition in telecommunications (2004)
8. Shapiro, C., Varian, H.R.: The art of standards wars. Managing in the modular age **41**, 247–272 (1999)

9. Katz, M.L., Shapiro, C.: Network externalities, companition and compatibility. Am. Econ. Rev. **75**, 424–440 (1985)
10. Shapiro, C., Varian, H.R.: Information rules: a strategic guide to the network economy. Harvard Bus. Sch. Press, Boston (1998)
11. Arthur, W.B.: Competing technologies, increasing returns, and lock-in by historical events. Econ. J. **89**, 116–131 (1989)
12. Eisenmann, T., Parker, G., Van Alstyne, M.W.: Strategies for two-sided markets. Harvard Bus. Rev. **84**, 92 (2006)
13. Hann, I.-H., Hui, K.-L., Lee, S.-Y.T., Png, I.P.: Overcoming online information privacy concerns: an information-processing theory approach. J. Manag. Inf. Syst. **24**, 13–42 (2007)
14. Laudon, K.C.: Markets and privacy. Commun. ACM **39**, 92–104 (1996)

The Conceptual Confusion Around "e-service": Practitioners' Conceptions

Eva Söderström[1(✉)], Jesper Holgersson[1], Beatrice Alenljung[1], Hannes Göbel[2], and Carina Hallqvist[2]

[1] School of Informatics, University of Skövde, Skövde, Sweden
{eva.soderstrom,jesper.holgersson,beatrice.alenljung}@his.se
[2] Section for Information Technology, University College of Borås, Borås, Sweden
{hannes.gobel,carina.hallqvist}@hb.se

Abstract. The e-service concept has been a central concern in many research and practitioner areas in recent years. There are expectations of citizens, customers, commercial companies and public organizations of what e-services are, their functionality and benefits. However, there is conceptual confusion that may hamper collaboration and research viability. This paper explores the conceptual vagueness and presents an empirical investigation of how the e-service concept is treated in practice, along with its kindred concept "IT service". Results show that public and commercial organizations approach e-services differently, that translation problems can cause lack of comparability in research results, and that additional concepts may be introduced instead of e-service.

1 Introduction

The need for providing services using information and communication technology (ICT) has multiplied concurrently with the growth and increased importance of ICT in society, for public administrations as well as business organizations. Nowadays, customers and citizens expect services to be electronically available, and the e-service concept has been a central concern in several research areas in recent years, e.g. e-business [1], IT development and maintenance [2], and e-government [3]. It has been described by practitioners as well as researchers, and quite a few researchers have tried to explain what "e-service" is [4, 5]. However, a universal definition is lacking, and "e-service" hence suffers from conceptual vagueness. The purpose of this paper is to investigate the conceptual vagueness and its consequences from a practitioner perspective, and to revitalize the conceptual discussion about e-services. In particular, the discussion will be conducted in relation to its kindred concept IT service.

2 Framing the Concepts e-service and IT Service

A "service" is traditionally seen as a set of activities provided by a provider to a consumer in order to generate value for both parties [6]. However, "service" is associated with a wide variety of meanings, not the least depending on the current context, and is thus

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 366–371, 2015.
DOI: 10.1007/978-3-319-25013-7_29

burdened with a clutter of meanings. There is no commonly agreed definition of *"e-service"* [4], but in a broad sense, e-service is seen as service delivered via electronic networks [4, 7]. Most research also agrees that e-services are based on interactivity and "driven by the customer and integrated with related organizational customer support processes and technologies" [8, p. 186]. It is a consumer who initiates interaction by requesting a service from an e-service provider. Researchers define e-services differently. For example, Javalgi et al. [7] says that e-services are interactive services delivered via the Internet whereas Rowley [4, p. 341] defines e-service as "deeds, efforts or performances whose delivery is mediated by information technology". Many researchers, however, take the concept for granted and do not define it at all. Instead, e-service is treated as something that is commonly known [e.g. 9]. Traditionally, IT has had a supporting role for businesses. By combining new technologies with the "new" service dominant logic paradigm [e.g. 10] new opportunities for service innovation emerge [e.g. 11]. For IT services, the field of IT Service Management (ITSM) is a key point of origin. ITSM is a widespread area where private and public sectors both have to manage and maintain IT-systems and processes as services. Within this field, Information Technology Infrastructure Library (ITIL) is one of the recognized large and extensive frameworks [12]. ITIL views an IT service as a service offered by an IT service provider. In contrast to the ITIL view of IT services, Jia and Reich [13] claim that the IT service concept traditionally has been described as a human mediated service delivered by IT personnel to business clients. This insinuates that an IT service is only related to the support provided to a user by a helpdesk function. It is also a more narrow view of the IT service concept than the one suggested by ITIL, thus emphasizing the conceptual confusion in the area.

3 Research Design

The study has a qualitative research approach in which the conceptual views and interpretations of e-service and IT service in different organizations have been investigated. Data was collected from 7 municipalities, 5 small and medium-sized IT enterprises (SMEs), and 1 regional alliance of municipalities. The interviewees were chosen based their potential to provide rich information concerning the concepts in focus.

3.1 Interviews

Open-ended interviews were conducted in which a semi-structured interview guide was used [14]. This ensured a solid basic part of the interviews, and gave flexibility to add questions when needed. The questions covered: (a) if some of the concepts e-service or IT service are used in the organization, (b) if other related similar concepts are used, (c) the interviewees perception of the concepts, and (d) if there are organization-collective definitions of the concepts. The interviews have been performed by various combinations of researchers, thus allowing for investigator triangulation [14]. Most interviews were conducted at the participant's workplace, some through the phone or email due to geographical distance. Each lasted for about 15–20 min, were recorded and subsequently transcribed.

3.2 Data Analysis

The qualitative data analysis was conducted in three steps, with an emphasis on researcher triangulation [14]. (1) Each researcher separately walked through the transcripts for their own perception of the material without being influenced by the others. (2) The researchers agreed to review the material from these dimensions: (a) similarities and differences within public organizations; (b) Similarities and differences within commercial organizations; and (c) similarities and differences between public and commercial organizations. (3) The researchers conducted a joint analysis using a whiteboard and color coding. Each respondent was given an identifier (letter + number): C for companies, M for municipalities, and LGF for the regional alliance.

4 Empirical Conceptual Elaboration

4.1 Similarities and Differences Within Public Organizations

In public organizations, the e-service concept is widely used and mostly referring to the same thing: services that previously were handled manually are now also offered via the Internet. The following quotation is an example of this view:

"*…E-service for me is something that is targeting citizens digitally*."(ME)

The focus is on citizens, but also companies. Public organizations often speak of citizens as external end users of the e-services, as illustrated by this quotation:

"*An e-service is […] a self-service that I can use to keep in contact with the municipality or a public authority, […] and that I can do it anytime. If the e-service is really good, I think it should have connection straight into the business systems so that it results in more efficiency*" (LGFA).

Accordingly, e-services need to provide value, for citizens and/or commercial organizations, for the municipality, or for both. This view is in contrast to a more general perspective in the e-government research community emphasizing that e-service mostly is provided by public administrations as a means to enhance internal efficiency. However, some municipalities claim that "real" e-services must provide value for both citizens and municipalities:

"*I do not think it is a real e-service when it is only the citizen that benefits, while the internal handling is the same as before. You spend the same amount of time*." (MC)

Mutual value is illustrated by the Swedish Association for Local Authorities and Regions, who say there is evidence that new and efficient e-services have contributed to reduce administrative costs for commercial organizations with 7 billion SEK, and that the e-services have reduced wrongful payments to citizens with 150 million SEK per year [15]. Some municipalities view digitized forms as e-services, while others view e-services as being those who cover an entire chain from citizens into the organization's ICT systems:

"*We are talking smart e-services […] that get into the various organizational systems*." (MD)

For some municipalities, mutual value is key while others are satisfied with increased value for only citizens. In contrast to "e-service", the majority of the municipalities do not use IT service at all, they simply state that they are not familiar with or are not using that as a concept in their organizations. Those who do use it or relate to it in some way view "IT service" as the internal IT department and helpdesk service:

> "*IT support is what we use, you say computer support or IT support but this is more practical. You want help with something concerning IT.*" (MF)

4.2 Similarities and Differences Within Commercial Organizations

In commercial organizations, an e-service can be defined in many different ways, but primarily connected to the Internet and to end-users and what they can do online:

> "*A traditional service that is accessible via a network-based interface, typically implemented using web technology. Preferably services provided by public authorities*." (CG)

The focus of companies seems to be on IT, and on service offerings using IT. In commercial organizations, "e-service" is not as prominent as "IT service". A common view of the IT-concept is in line with definitions provided by existing frameworks:

> "*Yes, the idea is that we must create value by managing the results that the client wants without the need to take ownership of specific "risks". IT Services is really this concept but applied to people and technology in an IT organization.*" (CE)

Hence, the definition of IT service is wide and does not focus only on end-users. One reason may be that existing frameworks such as ITIL are commonly used in commercial organizations, who therefore inherit the definitions used in the frameworks. It should be noted that not all commercial organizations use the term IT service, but rather have a plethora of service types that they discuss:

> "*I actually think we mostly use the service concept […] that is because we know what area we operate within and what area we focus on […] Well we know we work with IT services so perhaps that is why we do not define it so explicitly*." (CB)

Some companies differentiate between IT service and e-service in a different way, but referring to IT services as something internal and e-service as being external. Others, however, view e-services as being for organizational development instead. One very common view of e-services in commercial organizations is that the concept is associated with public authorities rather than commercial organizations:

> "*We don't use those concepts [e-service and IT service] in our organization, they are more used within the public sector*." (CF)

4.3 Similarities and Differences Between Public and Commercial Organizations

When merging material for the two organizational types, several similarities and differences can be identified. The ITIL framework, for example, colors the commercial

organizations' view of the service concept, which differs somewhat from how public authorities define it. One key aspect we identified is that of translation ambiguity of the concept "service" to other languages. Our research was conducted in a Swedish setting, and the Swedish language can translate "service" in two ways: One is focused on what is performed rather than on the technology mediating the service, while the other is technology-focused in terms of the technology used being the center of attention, such as in the ITIL definition of IT service. Commercial organizations are to some extent aware of the dual meaning of the "service" concept, while municipalities mainly refer to the service concept in relation to "support". This is natural since public organizations always have been focused on servicing their citizens and commercial organizations. The commercial organizations base and develop services focused on IT, involving IT technology, as well as processes and people that use the technology.

5 Concluding Analysis

Public and commercial organizations both differentiate between internal and external e-services. For example, CA expressed that e-services are services to end customers, which indicates that there are other services that are internal. People attach different meanings to concepts, and a common definition is often lacking:

"*We do not have a common definition [of the e-service concept] and we suffer from that.*" (CA).

A consequence of different meanings is that respondents may answer questions originating from one meaning, while researchers collecting data, or the collaborating organization had a different meaning in mind. The risk is that the interpretation does not represent the actual views of the other, which can make e.g. research results flawed and difficult to compare. Our study shows that there is a conceptual confusion based on both language and interpretation, and that definitions and scope vary within and between public and commercial organizations. Failing to ensure that collaborating partners, customers and providers, etc. mean the same thing can thus result in great problems. A common ground needs to be documented in any collaboration, in particular if translation is an issue. Our findings showed translation problems between Swedish and English, but this problem may hold true for other languages as well. Whether or not this is the case can only be established when a common ground is in place. Another dimension of the e-service conceptual discussion is what counts as a "real" e-service and what does not. Opinions vary, and even if research has discussed this issue to some extent, there is a difference with how it is discussed in public and commercial organizations. Future research should adopt a practitioner's perspective and conduct studies focused on empirical application of these levels. The purpose of this paper was to draw attention to the problem of conceptual vagueness and its consequences, and to revitalize the conceptual discussion about e-services. Our findings are a start of such a discussion, and future research needs to deepen and expand e-service research concerning its vagueness and confusion.

References

1. Janita, M.S., Miranda, F.J.: The antecedents of client loyalty in business-to-business (B2B) electronic marketplaces. Ind. Mark. Manag. **42**, 814–823 (2013)
2. Lu, J., et al.: Recommender system application developments: a survey. Decis. Support Syst. **74**, 12–32 (2015)
3. Fakhoury, R., Aubert, B.: Citizenship, trust, and behavioural intentions to use public e-services: the case of Lebanon. Int. J. Inf. Manag. **35**, 346–351 (2015)
4. Rowley, J.: An analysis of the e-service literature: towards a research agenda. Internet Res. Electron. Netw. Appl. Policy **16**(6), 879–897 (2006)
5. Rust, R.T., Lemon, K.N.: E-service and the consumer. Int. J. Electron. Commer. **5**(3), 85–101 (2001)
6. Grönroos, C.: Service logic revisited: who creates value? And who co-creates? Eur. Bus. Rev. **20**, 298–314 (2008)
7. Javalgi, R., Martin, C., Todd, P.: The export of e-services in the age of technology transformation: challenges and implications for international service providers. J. Serv. Mark. **18**(7), 560–573 (2004)
8. de Ruyter, K., Wetzels, M., Kleijnen, M.: Customer adoption of e-service: an experimental study. Int. J. Serv. Ind. Manag. **12**(2), 184–207 (2001)
9. van Velsen, L., et al.: Requirements engineering for e-Government services: a citizen-centric approach and case study. Gov. Inf. Q. **26**(3), 477–486 (2009)
10. Vargo, S.L., Lusch, R.F.: Evolving to a new dominant logic for marketing. J. Mark. **68**, 1–17 (2004)
11. Barret, M., Davidson, E., Prabhu, J., Vargo, S.L.: Service innovation in the digital age: key contribution and future directions. MIS Q. **39**(1), 135–154 (2015)
12. Cannon, D.: ITIL Service Strategy. The Stationary Office, Norwich (2011)
13. Jia, R., Reich, B.H.: IT service climate, antecedents and IT service quality outcomes: some initial evidence. J. Strateg. Inf. Syst. **22**(1), 51–69 (2013)
14. Patton, M.Q.: Qualitative Evaluation and Research Methods, 2nd edn, p. 532. Sage, Newbury Park (1990)
15. SKL, Strategy for the e-society (in Swedish), S.A.o.L.A.a. Regions, Editor (2011)

Social Customer Relationship Management: An Architectural Exploration of the Components

Marcel Rosenberger(✉)

Institute of Information Management, University of St. Gallen,
St. Gallen, Switzerland
marcel.rosenberger@student.unisg.ch

Abstract. In the recent years, social media have rapidly gained an increasing popularity. Companies have recognised this development and anticipate advantages from using social media for commercial purposes. Social customer relationship management (CRM) professionalises the use of social media and aims at integrating customers into operational procedures. This induces changes of existing structures, e.g. culture and organisation, business processes, information systems (IS), data structures, and technology. The intended transformation from CRM to social CRM is a complex task, because different aspects are affected, which also are mutually dependent. A prerequisite for the successful implementation of social CRM is understanding these aspects and its dependencies. Separation of concerns is a useful means of addressing complexity. The conglomeration of different issues is dissolved by conceptualising components and its relationships. This paper separates the concerns of social CRM using architectural perspectives and aims at building a better understanding. The research method is a literature review in which artefacts are gathered and assigned to five layers, which are business, process, integration, software, and technology. The conclusion states that social CRM is an emergent research field and comprises a call for more artefacts that concretise abstracted components of the business-layer.

Keywords: Social CRM · Design science · Enterprise architecture · Artefacts · Literature review

1 Introduction

Social media have gained interest and popularity in the past years. They are applications that build on web 2.0, which is a concept that encourages connecting, participation, and collaboration of users and sharing of content over the Internet [1]. The high number of social media users attracts attention of companies, which aim to profit from the potentialities [2, 3]. At a first glance, with only little effort companies can use social media to publish advertisements that reach many people, which improves the marketing efficiency. This view, however, is short-sighted. A closer study reveals more opportunities, which are enabled by the integration of social

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 372–385, 2015.
DOI: 10.1007/978-3-319-25013-7_30

media and the consumers into operational procedures. Examples of the potentialities are support-cost reduction, product innovation, and improvement of the reputation [4–6]. Social customer relationship management (CRM) is a philosophy and business strategy that professionalises the relationship to customers using social media and aims at realising the opportunities [7].

The transformation of an organisation from CRM to social CRM is a complex task, because many different aspects are affected, which also are mutually dependent. For example, Askool and Nakata [8] highlight that a strategy must be developed to govern social CRM initiatives. The management's task is to provide for a supportive company culture and implement organisational changes [9]. Existing information systems (IS) need adjustments to enable and enhance business processes. Finally, Social CRM requires integrations on functional and technological level [10–12]. Without understanding social CRM and its components, implementation projects are likely to fail. However, a holistic view of the components is still missing. The existing literature either focuses on single aspects or provides an abstracted overview of multiple aspects without giving details [13, 14]. This is justifiable in consideration of the complexity of social CRM. Still, a complete picture is desirable. Separation of concerns is a useful means of addressing complexity. The conglomeration of different issues is dissolved by conceptualising the components and its relationships. Artefacts, which are the results of design science, document components and its relationships of a domain of interest. They contribute to actual design-oriented business problems and support the implementation of technology-based solutions [15, 16]. This paper aims to answer the following research question (RQ).

RQ: What are the components of social CRM from an architectural perspective?

The intention is to build a better understanding of social CRM from an enterprise architect's view. Instead of proposing another abstracted framework or deep diving into parts of the complex, the present paper reuses existing research results and integrates the findings into a holistic view. The concerns of social CRM are separated using connected architectural perspectives. This allows investigating social CRM focused on specific aspects and in its entirety. The research method is a literature review, which allows determining the current state of research. The artefacts of the discovered publications are assigned to five layers, which represent the architectural perspectives business, process, integration, software, and technology. The layers are adopted from Enterprise Architecture (EA), which is a holistic framework that helps representing an enterprise's artefacts. Each layer is a view from the perspective of a specific concern. All artefacts of all layers represent the entire body of knowledge within the research scope. The target groups of this paper are IS architects and researchers of social CRM.

The paper is structured in five sections. Section 2 gives the background of the underlying concepts, which are social CRM, artefacts, and EA. The research method is described in Sect. 3. Then, the findings are presented considering each architectural layer. Finally, the paper concludes with a discussion and interpretation and guides further research.

2 Conceptual Foundation

Figure 1 provides an overview of the architectural exploration of the artefacts of social CRM. Each EA-layer may contain artefacts, and a single artefact may also address concerns of multiple layers mutually. Artefacts of social CRM and artefacts of CRM are included to broaden the perspective.

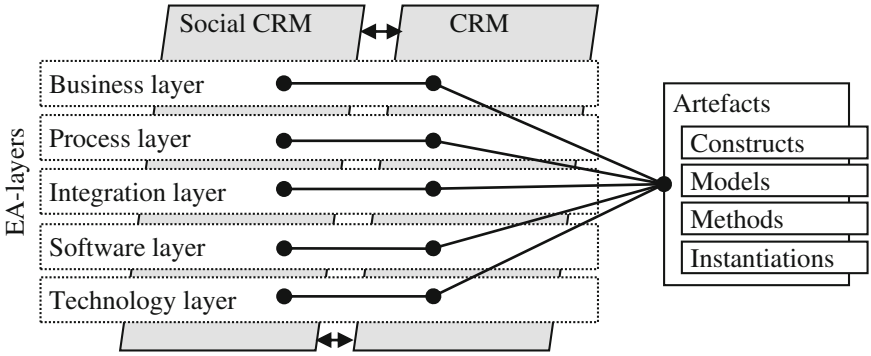


Fig. 1. Overview of the architectural exploration

2.1 Social CRM

Web 2.0 is a concept of the internet, which enables users to create content collaboratively and build a network with other users. The characteristics are user participation, openness, and network effects [1]. Openness means that results from user participation (e.g. posts, comments, and profiles) are accessible by other parties of the community. The concept is the foundation for social media, which are applications of the web 2.0 concept.

A basic feature of social media is to connect to other users to share information with them [17]. This principle is beneficial for the CRM of the company, which has the objective to establish and maintain profitable relationships to key customers and customer segments. CRM is a strategic approach that "involves identifying appropriate business and customer strategies, the acquisition and diffusion of customer knowledge, deciding appropriate segment granularity, managing the co-creation of customer value, developing integrated channel strategies and the intelligent use of data and technology solutions to create superior customer experience" [18].

Companies participate in the social network of users connecting to its target group. This facilitates the opportunity to gain business-relevant insights from the accessible data of the communication between the users. These insights help to intensify the relationship and to align the business with consumer needs. The integration of CRM with social media leads to the term "social CRM", which is a philosophy and business strategy [7]. Customers are engaged to participate in business processes with the result of a value-added for both: the company and the customer.

2.2 Artefacts

Design science research is a paradigm that aims at solving real-world problems by designing general solutions. It is a fundamental IS discipline, which develops artefacts that improve the capabilities of organisations [15]. Generality means that an artefact solves a class of problems instead of an individual problem of a single organisation. March and Smith [19] identify four artefact-types, which are constructs, models, methods, and instantiations.

Constructs are the basic language of concepts needed to describe phenomena. Models build on constructs and relate them with each other. Methods describe activities to meet specified targets. These forgoing artefacts can be instantiated in specific implementations representing the fourth artefact-type. The two main evaluation criteria are that artefacts are innovative and valuable [20]. Artefacts are ideal candidates to answer the research question, because they make components and its relationships explicit.

2.3 Enterprise Architecture

EA is a holistic framework, which provides views of an organisation's system from the perspective of specific concerns [21]. According to ISO/IEC/IEEE 42010:2011(E) [22] architecture is defined as the "fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution". The elements of a system can be related to five EA-layers, which are business-layer, process-layer, integration-layer, software-layer, and technology-layer [23].

The business-layer represents the strategy and subsumes organisational goals and success factors, products/services, targeted market segments, core competencies and strategic projects. The process-layer contains models to represent organisational units, business locations, business roles, business functions, metrics and service flows, for example. Applications and enterprise services are associated with the integration-layer. The software-layer contains software-components and data resources. Hardware units and network notes operate on the level of IT infrastructure (technology-layer). Relationships exist between components associated to the same layer and across layers. In the context of social CRM, sales and support are connected processes, for example, which both are associated to the process-layer. Social media are applications, which ultimately run on hardware. Thus, a connection between components of the integration-layer and components of the technology-layer exists. Aier et al. [21] identify the dissolution of information silos as an exemplary means of use of EA, which is the intention of use in this paper.

3 Method

The method for finding the existing artefacts is a literature review. Vom Brocke et al. [24] propose guidelines of a rigour process of literature reviews. They state that not only results should be presented, but, to allow replicability, also the approach. Table 1 characterises the conducted literature review following the

taxonomy proposed by Cooper [25]. The focus (1) is on existing constructs, models, methods, and instantiations that support the design, implementation and governance of social CRM. The goal (2) is to connect to existing knowledge to solve the research problem on a conceptual level (3).

The perspective (4) can be characterised as neutral representation, because the position is unbiased. Practitioners and researchers of social CRM are the target audience (5). The results are representative (6) for the IS community because prominent data sources have been queried.

Table 1. Taxonomy of the conducted literature review (borrowing from [25]).

Characteristic	**Categories**			
(1) focus	research outcomes	research methods	theories	applications
(2) goal	integration	criticism	central issues	
(3) organisation	historical	conceptual	methodological	
(4) perspective	neutral representation		espousal of position	
(5) audience	specialised scholars	general scholars	practitioners	general public
(6) coverage	exhaustive	exhaustive and selective	representative	central/pivotal

A keyword search in the databases of AISeL, EBSCO, Emerald, IEEE, JSTOR, ProQuest, and Web of Science in title (TI), topic (TO), abstract (AB), keyword (KW) and full text (TX) fields was applied. The first search-string was built to find specific design science results containing the term "social CRM" in particular. The total number of hits without duplicates was low (24). As a consequence, the search-string has been broadened. The second search-string includes design science results that consider CRM and also social media or web 2.0. This ensures the inclusion of research results that are applicable to the research scope whereas the term "social CRM" is not used. Only reviewed publications have been considered to ensure the level of quality. Duplicate publications of the two searches have been removed. The relevance of the distinct papers has been determined by reading the full texts. For example, publications that defined the term "CRM" as "component reference model" or "core reaction model" have been treated as not relevant. Only original publications written in English have been incorporated. The artefact-type of the found artefacts has been determined and the publication has been assigned to an architectural-layer. In cases where no unequivocal assignment could be made, multiple assignments of the same publication to all fitting layers have been made.

4 Findings

Table 2 shows the numerical results of the literature review. The two search-strings, which ultimately lead to the relevant publications, are the following.

Search-string (1): *"social crm" AND "design science"*

Search-string (2): *(crm OR "customer relationship management") AND ("web 2.0" OR "social media") AND "design science"*

The keyword "design science" proved to be eligible, because it allows an efficient and effective search. Prior searches with the keywords "architecture" or "integration" did not lead to noteworthy results. Applying the keyword "model" shows results, but this term is more often used in the context of quantitative research and signifies statistical models and does not lead to the sought architecture elements. Both terms "elements" and "components" are too broad and do not reduce the results sufficiently.

Table 2. Numerical results of the literature review

Data source	Search fields	Search string		Publications	
		(1)	(2)	Total[a]	Relevant
AISeL	TI, AB, KW, TX	21	38	53	12
EBSCO	TI, AB, KW, TX	2	53	53	5
Emerald	TI, AB, KW	--	1	1	--
IEEE	TI, AB, KW	1	4	5	2
JSTOR	TI, AB, KW, TX	--	5	5	--
ProQuest	TI, AB, KW	3	46	46	1
Web of Science	TI, TO	--	--	--	--
Total[a]		**24**	**137**	**151**	**21**

[a]The total numbers are not equal to the column and row sums respectively, because duplicates have been counted only once.

Applying the search-string (1) to the data source AISeL displays 21 results. This indicates two different things. Firstly, social CRM is a present research field of the IS community. Secondly, design science is a common research paradigm of this research field. Apart from that, Emerald, JSTOR, and Web of Science display no results for search-string (1). A possible reason is that Emerald and JSTOR include mainly journal publications and no conference papers. The publication-period of conference proceedings is usually shorter. Hence, "social CRM" is a novel term that is not yet established in journals. The fact that JSTOR has results for "design science", but no results for "social CRM", supports that argument. Emerald does not feature to search the full texts of publications and the occurrence of "social CRM" and "design science" in the metadata

is non-existent. This also applies to Web of Science, where only title (TI) and topic (TO) fields are searchable. In total, search-string (1) leads to 24 unique publications.

Applying search-string (2) displays results in all data sources, except Web of Science due to the limitation to search in the abstracts or the full texts of the publications. The high total number of unique results (137) indicates that CRM is better established in research than social CRM. More precisely, the term "social CRM" is not as widely used as the combination of CRM and web 2.0 or social media. This is not surprising, because all three terms CRM, social media, and web 2.0 are the foundation and a prerequisite to define social CRM. EBSCO, ProQuest, and AISeL are the data sources with the highest totals for search-string (2).

In summary, 21 relevant publications describe artefacts that represent components of social CRM. Figure 2 shows the yearly distribution of the publications. The findings indicate that social CRM is a contemporary research field. All artefacts have been published in the past six years. The year 2012 is remarkable, because the number of publications increased threefold compared to the previous year. Since then, the number of publications that include artefacts relevant to social CRM amounts five to six per year in the queried data sources.

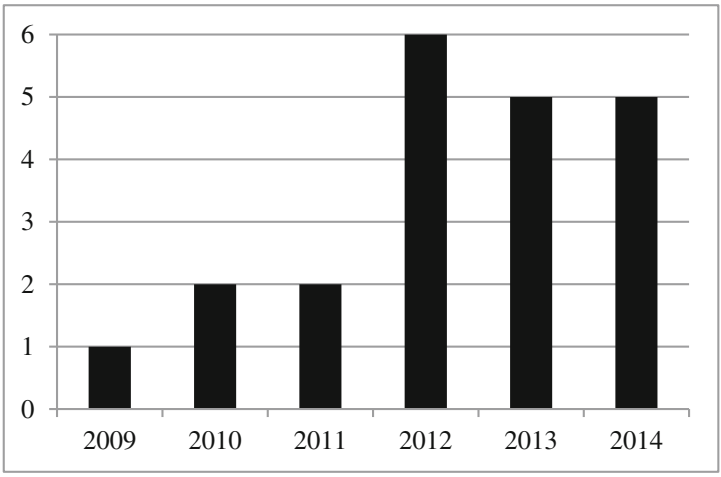


Fig. 2. Distribution of the relevant publications per year

Table 3 lists the publications, the artefact-types, the artefact-names, and the addressed EA-layers. A model is the most commonly occurring artefact-type among the explored publications. Eleven publications present models, five describe instantiations, three develop a method, and two publications propose constructs. Models, methods, and instantiations are built on constructs [19]. Consequently, every publication that describes an artefact contains or relies on constructs. In cases where models, methods, and instantiations implicitly rely on artefacts of other types, only the targeted artefact of the publication is incorporated. As a result, the number of counted constructs can be less than the number of the other artefact-types.

The business-layer includes the most artefacts. Ten artefacts address aspects of the business-layer, nine thematise processual issues, nine target applications of social CRM (integration-layer), seven outline software issues, and no artefact depicts components of the IT infrastructure (technology-layer). The artefacts of ten publications address concerns of multiple EA-layers. Examples are framework-models, which provide different views of components and its relationships [13, 26, 27]. Five artefacts address aspects of the integration-layer and software-layer in conjunction. The combined consideration of the business-layer and process-layer occurs four times and is second most therewith. This indicates a close architectural relationship between these layers.

A model of the business-layer is most frequent counting seven occurrences. Four publications develop models that highlight processual issues and four other publications describe instantiations that relate to the integration-layer implementing applications.

Table 3. Explored publications, artefact-types, artefact-names, and addressed EA-layers

	EA-layer					**Artefact-type**				
Publi-cation	Business	Process	Integration	Software	Technology	Constructs	Model	Method	Instantiation	**Artefact**
[28]	○	○	●	●	○	○	●	○	○	*Framework of an ontology-based social media analysis*
[29]	○	●	○	○	○	○	○	○	●	*Text mining application for exploring the voice of the customer*
[30]	○	○	○	●	○	○	●	○	○	*Data model of data objects of social networks*
[31]	○	○	●	○	○	○	●	○	○	*Framework for gathering business intelligence from blogs*
[32]	○	●	●	●	○	○	○	○	●	*Social CRM tool framework*
[33]	●	○	○	○	○	○	○	●	○	*Method for developing a taxonomy of social media*
[14]	○	○	●	●	○	○	○	○	●	*User interface design for Twitter*
[34]	●	●	○	○	○	○	○	●	○	*Social media innovation method*
[35]	○	○	●	○	○	○	○	○	●	*Social data analytics tool (SODATO)*
[26]	●	●	○	○	○	○	●	○	○	*Social CRM framework*
[36]	●	○	○	○	○	●	○	○	○	*Web 2.0 factors and value drivers*
[37]	●	●	●	○	○	○	●	○	○	*Maturity model for the adoption of social media*
[38]	●	●	○	○	○	○	●	○	○	*System dynamics model and word-of-mouth effects*
[39]	○	●	○	○	○	●	○	○	○	*Ontology for IS sentiment analysis*
[13]	●	○	○	○	○	○	●	○	○	*Enterprise 2.0 management framework*
[40]	○	●	●	●	○	○	●	○	○	*Blueprint of an analytical social CRM system*
[41]	○	○	○	●	○	○	●	○	○	*Social network data model*
[42]	●	○	○	○	○	○	●	○	○	*Customer satisfaction theoretical framework*
[43]	○	○	●	●	○	○	○	○	●	*Multimedia platform providing social e-services*
[27]	●	○	○	○	○	○	●	○	○	*Social media strategy framework*
[44]	●	○	●	○	○	○	○	●	○	*Social app prototypes*

Legend: ○ not a focus of the publication; ● focus of the publication

4.1 Business-Layer

The artefacts of the business-layer address abstracted concepts, strategic aspects, organisational goals or success factors relevant to social CRM. Emamjome et al. [33] propose a method for developing a taxonomy of social media in an organisational context. The findings help to create an understanding of the concepts and support building a foundation for further research. Patten and Keane [13] conceive social CRM as a combination of three dimensions of a company-wide concept. Their Enterprise 2.0 Management Framework relates the dimensions (1) technology, tools and capabilities, (2) user-generated content, and (3) employee and customer applications, which are governed by culture and policies.

The maturity model of organisational adoption of social media allows positioning and comparing of the companies' achievements along the dimensions strategy, processes, IS, culture, and governance [37]. The artefact is useful to determine and adjust the approach to social media from a strategic perspective. Factors and value drivers of web 2.0 support the creation of a social CRM strategy, because they show causes and effects in a business context. Lehmkuhl and Jung [36] identify social networking, interaction orientation, user-added value, and customisation/personalisation as factors, which have a varying relevance for a specific company depending on the business model. A commerce-oriented online marketplace, for example, should encourage the customer-company and customer-customer communication (interaction orientation). A content-oriented online newspaper can profit from gathering and exchanging information and opinions of its readers by enabling social networking. According to Werder et al. [27], the social media strategy should include three components, which are scope, capabilities and governance. The scope is defined by actors, platforms and interaction. Social media objectives and activities are conceived as capabilities. The governance-component addresses value, resources and risks. Customer experience and customer satisfaction are further strategic focus areas that need attention and allocation of funds [42].

Yoon et al. [44] develop a conceptualisation of social commerce identifying the components user contribution, participation, collaboration and technological features. By reference to the customer life cycle model they argue that web 2.0 supports business goals. Interesting to note, social commerce addresses similar aspects of what Greenberg [7] terms social CRM. A clear differentiation between both terms is missing. In summary, the artefacts of the business-layer help to understand basic concepts, identify strategic aspects for planning and organising social CRM, include recommendations for governance, and "[stimulate] thinking about the impact of social media beyond the marketing function" [27].

4.2 Process-Layer

In order to implement social CRM in organisations the management needs to introduce, adjust, and evaluate business processes. A framework model helps to scope the tasks and structure work packages. The core processes for planning social CRM are readiness assessment, strategy development, value creation, multichannel management, information management and performance measurement [26]. The implementation-activities

are governed by project management and change management and employee engagement. Helms et al. [34] focus on user participation in the innovation process developing a social media method for matching innovation tasks with social media characteristics. The characteristics are organised in three dimensions, which are audience, content and time. Botzenhardt et al. [29] focus on supporting the product development process. Their instantiation is a text mining software that analyses the unstructured content of customers' posts in social media. Maier and Reinwald [38] support the decision-making in the complaint management process. The authors propose a system dynamics model and incorporate the influence of social media on word-of-mouth effects and the customers' repurchase behaviour. Online social networks act as an accelerator and can have both: positive and negative effects.

The artefacts of the process-layer represent only some processes of social CRM. Other relevant processes are not covered, e.g. customer support, lead management, up-selling/cross-selling, and market research.

4.3 Integration-Layer

The artefacts of the integration-layer propose models of applications and describe real-life instantiations of social CRM applications. The framework of an ontology-based social media analysis is a social CRM system model [28]. The central component is the Text Mining Framework, which has social media interfaces to access data of social media and database interfaces to enrich the data of enterprise systems (e.g. CRM system). The Ontology Engineering component extracts domain concepts of the company (i.e. ontology) from the data of its enterprise systems. The extracted company information and information about products is forwarded to the text mining application, which uses the ontology for filtering relevant and irrelevant social media data. The insights from social media data can be used to enhance products or plan marketing campaigns, for example. Chau and Xu [31] propose a framework for collecting and analysing business intelligence in blogs. This model is more concrete and regards blogs, which are a type of social media. The approach is the same: content analysis techniques are used to gather insights from the user-created content. Deng et al. [32] apply network analysis technologies to enhance marketing and sales processes and implement an application. The publication contains details of the data resources and software components and thus is additionally assigned to the software-layer. The Social Data Analytics Tool is a software-instantiation that fetches the social graph and the social text from social media [35]. The social graph represents actors and actions. Sentiments, keywords, and topics are extracted from the social text. Spagnoletti and Resca [43] implement a multimedia online platform and highlight that an online-community is a valuable tool for CRM.

Most artefacts of the integration-layer concentrate on gathering insights from social media data using analytical technologies. The represented integrations between enterprise systems and social media have in common that they follow an extract-transform-load (ETL) approach. This leads to a unidirectional connection from social media to the enterprise systems. However, a communication needs a two-way integration.

4.4 Software-Layer

The artefacts of the software-layer model components of social CRM applications. Examples of components are the user interface and the data model [14]. Rosemann et al. [41] include social media data in a Business Intelligence (BI) system that is capable to report on characteristics, needs, wishes and demands of customers. These insights are extracted by analytical operations on a combined set of the data of the CRM system, the data warehouse and social media. Key activities and components of the analytical social CRM system can be organised by the groups social web, multi-channel-management, analytical CRM, data, and operational CRM [40]. Analytical operations on the data are analysis, reporting, monitoring, and generating. The operational activities are planning, executing, and controlling CRM processes.

The identified artefacts are valuable to describe the structure and function of social CRM applications. The models and instantiations either focus on single components of a social CRM application in detail or give an overview of multiple components and its relationships.

4.5 Technology-Layer

No artefact represents technological aspects of IT infrastructure, such as hardware units, network nodes, and physical servers. It is undisputed that technology is an important component of a social CRM system [7]. However, the artefacts discovered in the literature review do not provide details about necessary or recommended IT infrastructure components.

5 Conclusion

Social CRM is an emergent research field with an increasing number of publications that present artefacts. However, the term is not commonly established. Besides “social CRM”, researchers use the terms “Enterprise 2.0”, “CRM 2.0”, and “social commerce” without clear differentiation [13, 44, 45]. The findings help to establish a better understanding of social CRM, because the explored artefacts reveal the components and its relationships. Social CRM, in entirety, comprises aspects of the business-, process-, integration-, software-, and technology-layers. No single artefact, however, covers all components of all layers mutually. This is explainable by the complexity of social CRM and, in consequence, the need to examine the philosophy and business strategy on different levels of abstraction and by separation of concerns. No discovered artefact represents components of the technology-layer. A possible reason is that the setup of hardware units, network nodes and servers are specific to an organisation and depend on company-size and individual organisational requirements. However, research results should also be applicable to other (similar) situations. Artefacts provide a general solution to a problem in a specified context [15, 19]. Hence, it is not surprising that the technology-layer is under-represented in research. On the contrary, artefacts of the technology-layer are a possible research output. Examples are design principles of successful infrastructure-setup to accommodate the high load of the analytical social media data processing.

The findings have managerial impacts. Not all artefacts, however, are relevant to all stakeholders of social CRM. For example, the management of an organisation might want to adduce artefacts of the business-layer to refine the social media strategy, while developers receive conceptual guidance from the models and instantiations of the integration-layer and software-layer in particular.

A limitation of this research is that the quality of the discovered artefacts has not been evaluated. Furthermore, due to the research question, only design science results, i.e. artefacts, are included in the literature review.

Further research may focus on components of a single layer, multiple layers, or the connection of the layers. A higher layer gives orientation and determines aspects, which need further concretisation on a lower layer. For example, culture and governance are mentioned as important issues of the business-layer [13, 37]. However, no artefact on the process-layer continues these aspects. It is not sufficient to identify *what* the important components of social CRM are, but also *how* the components function. In summary, the findings lead to a call for more artefacts that concretise the components of the business-layer. Especially constructs, which define a common terminology and methods that guide the successful implementation of social CRM components are sparse.

References

1. Musser, J., O'Reilly, T.: Web 2.0 principles and best practices (2006)
2. Smith, T.: Conference notes – the social media revolution. Int. J. Mark. Res. **51**, 559 (2009)
3. Baird, C.H., Parasnis, G.: From social media to social customer relationship management. Strateg. Leadersh. **39**, 30–37 (2011)
4. Cappuccio, S., Kulkarni, S., Sohail, M., Haider, M., Wang, X.: Social CRM for SMEs: current tools and strategy. In: Khachidze, V., Wang, T., Siddiqui, S., Liu, V., Cappuccio, S., Lim, A. (eds.) iCETS 2012. CCIS, vol. 332, pp. 422–435. Springer, Heidelberg (2012)
5. Fliess, S., Nesper, J.: Understanding patterns of customer engagement – how companies can gain a surplus from a social phenomenon. J. Mark. Dev. Competitiveness **6**, 81–93 (2012)
6. Jahn, B., Kunz, W.: How to transform consumers into fans of your brand. J. Serv. Manag. **23**, 344–361 (2012)
7. Greenberg, P.: CRM at the Speed of Light: Social CRM 2.0 Strategies, Tools, and Techniques for Engaging your Customers. McGraw-Hill Osborne Media, New York (2010)
8. Askool, S., Nakata, K.: A conceptual model for acceptance of social CRM systems based on a scoping study. Ai Soc. **26**, 205–220 (2010)
9. Kuikka, M., Äkkinen, M.: Determining the challenges of organizational social media adoption and use. In: ECIS 2011 Proceedings, Paper 248 (2011)
10. Acker, O., Gröne, F., Akkad, F., Pötscher, F., Yazbek, R.: Social CRM: how companies can link into the social web of consumers. J. Direct Data Digit. Mark. Pract. **13**, 3–10 (2011)
11. Reinhold, O., Alt, R.: How Companies are implementing social customer relationship management: insights from two case studies. In: Proceedings of the 26th Bled eConference, pp. 206–221 (2013)
12. Trainor, K.J., Andzulis, J.M., Rapp, A., Agnihotri, R.: Social media technology usage and customer relationship performance: a capabilities-based examination of social CRM. J. Bus. Res. **67**, 1201–1208 (2013)
13. Patten, K., Keane, L.: Enterprise 2.0 management and social issues. In: AMCIS 2010 Proceedings, Paper 395 (2010)

14. Gruzd, A.: Emotions in the twitterverse and implications for user interface design. AIS Trans. Hum. Comput. Interact. **5**, 42–56 (2013)
15. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**, 75–105 (2004)
16. Müller, B., Olbrich, S.: The artifact's theory – a grounded theory perspective on design science research, pp. 1176–1186 (2011)
17. Ang, L.: Is SCRM really a good social media strategy? J. Database Mark. Cust. Strateg. Manag. **18**, 149–153 (2011)
18. Frow, P.E., Payne, A.F.: Customer relationship management: a strategic perspective. J. Bus. Mark. Manag. **3**, 7–27 (2009)
19. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support Syst. **15**, 251–266 (1995)
20. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**, 45–77 (2007)
21. Aier, S., Gleichauf, B., Winter, R.: Understanding enterprise architecture management design – an empirical analysis, pp. 645–654 (2011)
22. ISO/IEC/IEEE: Systems and software engineering - Architecture description, Reference number ISO/IEC/IEEE 42010:2011(E). IEEE Comput. Soc. (2011)
23. Winter, R., Fischer, R.: Essential layers, artifacts, and dependencies of enterprise architecture. In: Proceedings of 2006 10th IEEE Int. Enterprise Distributed Object Computing Conference Work, EDOCW2006, pp. 1–12 (2006)
24. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., Brocke, J. Von, Reimer, K.: Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: 17th European Conference on Information Systems (2009)
25. Cooper, H.M.: Organizing knowledge syntheses: a taxonomy of literature reviews. Knowl. Soc. **1**, 104–126 (1988)
26. Lehmkuhl, T.: Towards social CRM - a model for deploying web 2.0 in customer relationship management (2014)
27. Werder, K., Helms, R., Slinger, J.: Social media for success: a strategic framework (2014)
28. Alt R., Wittwer M.: Towards an ontology-based approach for social media analysis. In: Proceedings 22nd European Conference on Information Systems. Tel Aviv, pp. 1–10 (2014)
29. Botzenhardt, A., Witt, A., Maedche, A.: A text mining application for exploring the voice of the customer (2011)
30. Braun, R., Esswein, W.: Corporate risks in social networks–towards a risk management framework. In: 18th American Conference on Information Systems, pp. 1–12 (2012)
31. Chau, M., Xu, J.: Business intelligence in blogs: understanding consumer interactions and communities. MIS Q. Manag. Inf. Syst. **36**, 1189–1216 (2012)
32. Deng, X.L., Zhang, L., Wang, B., Wu, B.: Implementation and research of social CRM tool in mobile BOSS based on complex network. In: 2009 1st International Conference on Information Science and Engineering, ICISE 2009. pp. 870–873 (2009)
33. Emamjome, F., Gable, G.G., Bandara, W., Rabaa'i, A.: Understanding the value of social media in organisations: a taxonomic approach. In: PACIS 2014 Proceedings, Paper 59 (2014)
34. Helms, R.W., Booij, E., Spruit, M.: Reaching out: involving users in innovation tasks through social media. In: ECIS 2012 Proceedings, Paper 193 (2012)
35. Hussain, A., Vatrapu, R.: Social data analytics tool (SODATO). In: Tremblay, M.C., VanderMeer, D., Rothenberger, M., Gupta, A., Yoon, V. (eds.) DESRIST 2014. LNCS, vol. 8463, pp. 368–372. Springer, Heidelberg (2014)

36. Lehmkuhl, T., Jung, R.: Value creation potential of web 2.0 for Sme — insights and lessons learnt from a European producer of consumer electronics. Int. J. Coop. Inf. Syst. **22**, 1340003 (2013)
37. Lehmkuhl, T., Baumöl, U., Jung, R.: Towards a maturity model for the adoption of Social Media as a means of organizational innovation. In: Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 3067–3076 (2013)
38. Maier, M., Reinwald, D.: A system dynamics approach to value-based complaint management including repurchase behavior and word of mouth. In: ECIS 2010 Proceedings, Paper 55 (2010)
39. Park, E., Storey, V., Givens, S.: An ontology artifact for information systems sentiment analysis (2013)
40. Reinhold, O., Alt, R.: Analytical social CRM: concept and tool support. In: Proceedings 24th Bled eConference, pp. 226–241 (2011)
41. Rosemann, M., Eggert, M., Voigt, M., Beverungen, D.: Leveraging social network data for analytics CRM strategies - the introduction of social BI. In: ECIS 2012 Proceedings (2012)
42. Seng, W.M.: E-government evaluation: the use of design science approach to build and evaluate customer satisfaction theoretical framework. In: 2012 Seventh International Conference on Digital Information Management (ICDIM), pp. 266–273 (2012)
43. Spagnoletti, P., Resca, A.: A design theory for IT supporting online communities. In: 2012 45th Hawaii International Conference on System Sciences, pp. 4082–4091 (2012)
44. Yoon, S.Y., Studies, P., Technologies, E.: Empirical investigation of Web 2.0 technologies for social commerce and implementation of social app prototypes (2013)
45. Trainor, K.J.: Relating social media technologies to performance: a capabilities-based perspective. J. Pers. Sell. Sales Manag. **32**, 317–331 (2012)

Removing the Blinkers: What a Process View Learns About G2G Information Systems in Flanders (Part 2)

Lies Van Cauter[1(✉)], Monique Snoeck[2], and Joep Crompvoets[1]

[1] KU Leuven, Public Governance Insitute, Leuven, Belgium
{Lies.VanCauter,Joep.Crompvoets}@soc.kuleuven.be
[2] KU Leuven, Research Centre for Management Informatics, Leuven, Belgium
Monique.Snoeck@kuleuven.be

Abstract. Information sharing across the public sector is a precondition for innovation. The reality today is that data are scattered throughout administrative services. Creating government-to-government (G2G) information systems (IS) has the potential to sustain fluent data flows. Despite this potential, G2G IS projects fail to deliver the expected benefits. Factor research partially explains why so many G2G information systems fail. In this paper we take a broader perspective by applying process research to study six recurrent problems of Flemish G2G IS in their dynamic context. We test whether Sauer's needs and support-power analysis can provide additional management insights concerning G2G IS projects. Our results, based on interviews and focus groups, show that seemingly controllable problems have much deeper roots that require managers' action.

Keywords: IS failure · G2G · Process management

1 Introduction

In November 2014 the Organisation for Economic Co-operation and Development (OECD) called for action to enable public sector innovation. A core precondition is free flowing data, since shared information provides a basis for simplification, accountability or collaboration and allows organisations to learn collectively [17].

The potential added-value of free flowing data is high, but unfortunately, the reality today is that data are scattered throughout administrative services. The Weberian bureaucracy, characterized by its strict task allocation and hierarchy, has led to fragmentation of policy and service delivery. This problem is pervasive: the need for information sharing exists both across different levels of governments (vertical dimension) as among different governmental agencies (horizontal dimension).

ICT is perceived as an important driver of change because the creation of digital government-to-government (G2G) information systems (IS) has the potential to sustain free data flows [2]. Unfortunately G2G IS projects continue to fail to deliver expected benefits [4, 24]. This problem has heavily been researched during the last decades, without resulting in a great improvement of failure rates. One cause might be that for a long time a rational and technical view on failure dominated [19]. Positivistic researchers

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 386–397, 2015.
DOI: 10.1007/978-3-319-25013-7_31

and project managers believed that problems can be eliminated if failure factors are listed and if management can detect and eradicate these linear failure factors. Yet eliminating failure factors does not warrant success. Rational factor research ignores the context of an IS as well as the dynamic non-linear interactions of (non-)technical factors such as legislation, politics, economic or cultural factors [24]. Rational project managers tend to follow fixed goals and try to minimise the risk of random context events, as such they only see a part of the IS failure puzzle [12].

In 2014 Dwivedi et al. [8] called for research that includes a larger part of the failure puzzle by incorporating local contingencies and the dynamic environment of ISs. A growing research stream that connects to this call is the 'process perspective' (e.g. Lee and Liebenau, Markus and Robey, Sauer [8]). Process managers look at the interaction between an IS, its stakeholders and context factors. An IS project does not exist in a vacuum [12, 25], compatibility with a given environment is a key precondition for innovation [5]. Stakeholders of G2G ISs interact dynamically and may have diverse interests which can e.g. result in sabotage of project goals. A major cause of failure is the inability to deal with these [2]. The ability to adapt to environmental developments and changing stakeholders' needs, determines governmental innovation [2, 12]. Sauer believes that managers who are confronted with troubled ISs, could start with an analysis of their situation by conducting a needs assessment: What problems need to be solved and what stakeholder support would meet these. This should be followed by a support power analysis: Who has the power to provide the required support? By conducting such analysis, managers will be more aware about context and dynamic stakeholder interactions that influence their ISs (i.e. a process perspective) [22]. We elaborate further on Sauer's work, as the main research question is:

> *Can the needs and support-power analysis of Sauer provide additional insights for G2G IS management in Flanders?*

In a previous article [23] we conducted a needs and support-power analysis, but limited ourselves to the study of recurrent technological and political problems of G2G IS projects in Flanders. The analysis showed that factor research can appoint recurrent problems but that Sauer's process perspective provides extra insights concerning the context of these problems and support (difficulties) of relevant stakeholders. This article investigates whether conducting the same analysis for recurrent economic and juridical problems of G2G IS projects in Flanders, can add insights to the prior analysis results. The specific research question of the paper is: *Does the analysis of the economic and juridical recurrent problems, via the needs and support-power analysis of Sauer, provide additional insights for G2G IS management in Flanders?*

The structure of this paper is as follows. The theoretical framework is sketched in Sect. 2. Methodology is described in Sect. 3. The actual analysis is presented in Sect. 4, it is followed by a discussion of the results in Sect. 5. We conclude in Sect. 6.

2 Theoretical Framework

Sauer sees ISs as the product of a process which is open to flaws. This process consists of an initiation, development, implementation and operational phase, it may be problematic.

An innovation process can be split up in a need and a support management process which may be influenced by contextual uncertainties. In order to continue the innovation process, a project organisation requires enough support: support is searched during the support management process. If there is too little support, the endurance of the innovation process and the whole IS is threatened [22].

Sauer modelled a triangle of dependences: ISs exist to serve stakeholder's interests. They require a variety of support if they are to function at all. The IS's project organisation has a special role in innovating the system. Support for carrying out this role will only be given if supporters' interests are served. Managers could use a needs analysis to define the problems with which the project organisation will be confronted during the innovation process. This analysis further pictures the context influences and available problem solving mechanisms. A support-power analysis helps additionally to determine who has the potential power to provide the required support.

2.1 Needs Analysis

The project organisation's needs analysis will consist of two parts: (a) an analysis of problems and (b) an analysis of the required support to solve these problems:

a. Analysis of problems is twofold: (1) map the problems to be solved, (2) do a context scanning. Context helps to define problems but constraints originating in this context may make the innovation process problematic. The context is analysed along six dimensions: 1. human factors, 2. history, 3. technological process, 4. structure, 5. politics and 6. environment. Environment is subdivided in: 6.1. customers, 6.2. suppliers, 6.3. competitors, 6.4. technology, 6.5. regulators, 6.6. interests and 6.7. culture [22].
b. The analysis of support looks at available problem-solving mechanisms G2G ISs are often ineffective in transferring data between organisations. These systems are confronted with complex combinations of problems. Managers should understand problems in depth, in order to deal with these in an effective manner. In the project organisation the idea champion takes up this vital task. The potential of an innovation is also dependent on the context in which information sharing takes place [22].

2.2 Support Power Analysis

The support power analysis investigates who is able to provide support identified in the needs analysis and what other relations may affect stakeholders [22]. It can be applied at any stage of a G2G IS project. Sauer advices to conduct this analysis often. An idea champion should acknowledge his dependence on others. Information sharing requires both thinking about the own organisation and of external actors [17]. The latter may support/obstruct a project while trying to protect their core values and may react in an unpredictable way to interventions. Costs- benefits are not evenly divided, some stakeholders win, some lose. Idea champions should be sensitive for chances of random decision making, managing these becomes in itself dynamic, management strategies are dependent on the situation at hand and stakeholders' reactions [7, 12].

3 Methodology

3.1 Data Collection

In 2012 an exploratory research on trends and challenges of G2G IS projects was conducted. 20 experts of all Belgian governmental levels (i.e. local, provincial, regional and federal) were interviewed on this matter. In 2014 32 idea champions of G2G ISs in Flanders (Belgium) were brought together in five focus groups to discuss IS challenges, trends and the management thereof. We detect an overlap between the findings of both studies: managers in G2G IS projects face recurrent problems, which we structured via Sauer's 'needs and support-power analysis framework'. Technical/political problems were discussed in a previous paper [23]. This paper analyses recurrent economic/juridical problems and focuses primarily on vertical ISs between the Flemish regional government and local governments. However, since for some IS projects several Flemish organisations were involved, this adds a horizontal dimension. The next paragraphs describe data collection techniques in more detail.

Interviews 2012. Interviewing is a common data collection technique in IS research. In 2012, 20 experts on G2G IS projects were interviewed face-to-face on trends and challenges of G2G IS projects. Interviews were carried out over a three-month period. The interviewees represented local or provincial stakeholder groups, managed a successful G2G IS project or tried to monitor several G2G ISs. They worked for 14 different organisations at all governmental levels, this was a deliberate choice: by collecting different points of view, the risk of attribution bias was reduced. The interviewer asked open-ended questions to probe interviewees when interesting topics surfaced [20]. Both (non-)verbal language was captured. Each interview lasted between one and two hours and was transcribed with permission. The policy documents and legislation interviewees referred to, were studied as well. All interviewees received an end report. After 20 interviews, a point of saturation was reached.

Focus Groups 2014. A focus group (FG) is a group of individuals assembled by researchers to discuss and comment on a certain topic. It allows to obtain a variety of perspectives from a single data-gathering session [20]. Mainly the last decade FGs are gaining visibility and acceptance in IS research [3]. In 2012 we created an inventory of existing G2G ISs in Flanders. Based on this inventory, 40 IS idea champions were invited to participate in FG discussions. Five refused cooperation, five others did not show up. We slightly over-recruited the number of idea champions and reached as such the optimal number of 6 or 7 participants per session. A pretested questioning route was used to guide the conversations. The moderator briefed the participants, tried to create an informal sphere and ensured that everyone could have a say. She encouraged dialoguing via follow-up questions and by showing a stimulating body language. A senior researcher took up the role of assistant moderator. She observed body language, took notes and summarised the viewpoints. After 5 sessions saturation was achieved [3]. For a more detailed description of the data collection see [23].

3.2 Data Analysis Method

The interview and FG questions are not based on a specific theoretical model so that the data could speak for itself (an inductive approach). The five stages model of Krueger [14] was used to interpret the data. (1) Familiarisation: The researcher gets familiar with the major themes by reading the transcripts. (2) Themes: She develops categories within the major themes based on a questioning route. (3) Indexing: Data within and between cases are compared. (4) Charting: Data are reduced, important quotes rearranged under new codes. (5) Mapping and interpretation: Links between quotes are interpreted to make sense of the data as a whole. By studying the data, a series of G2G IS problems surfaced. We grouped these in 4 main categories: (A) technological, (B) political, (C) economic and (D) juridical recurrent problems. (A) Recurrent technological problems have to do with the business case, IT infrastructure, developers, planning and security. (B) Political recurrent problems involve top management support, user involvement and the skills/position of the idea champion. (C) Economic problems are about the need for (in)tangible resources. (D) Finally too much/less change in legislation, involuntary use and privacy form recurrent juridical problems. 287 pages of transcripts were coded in the qualitative data analysis programme NVivo. Data were analysed in two stages. (1) The problems detected from the interview data of 2012 were compared to the FG results of 2014. (2) Recurrent problems were compared to Sauer's framework by applying a needs and support analysis.

4 Analysis

The data analysis of interviews and FGs reveals an overlap in problems concerning political, technological, economic and juridical issues. Considering that these problems reoccur in both studies, we assume that they are rather structural and widespread for Flemish G2G IS projects. Factor research lists these problems too but ignores their interaction and context. It misses as such a part of the IS failure puzzle. In contrast, we take a process view by conducting a needs and support-power analysis on the recurrent problems. This enables us to research whether additional insights can be found via the process perspective. We found that the need and support analysis of political and technological recurrent problems indeed provides additional insights [23]. In the next paragraphs we will analyse recurrent economic and juridical issues and research whether these too provide additional insights. Due to space limits only the six most prominent economic and juridical recurrent problems are presented. Per problem three main things are described: (1) the problem to be solved (= problem description), (2) which context elements influence the problem (= context), and (3) which mechanisms can solve these problems and whose support is relevant therefor (= support). Every problem is influenced by several context elements. These elements are numbered and the applicable context category is mentioned between brackets. These numbers are referred to in the description of the support in order to motivate which support element relates to which element of the context.

4.1 Economic Agreements

To set up and run a G2G IS project, stakeholders must agree to provide tangible resources such as money/personnel and intangible resources (i.e. data, information) [4].

Problem 1: Money. *Problem Description.* A sufficient amount of money should be spent on a G2G IS for development, maintenance and adaptions: Who will finance what and when?

Context. Several context factors influence this need. (1) A macro factor is the economic crisis, due to budget cuts the willingness to do something for another government dropped (environment) and (2) funding to stimulate information sharing is under pressure (environment). (3) The budgetary capacity of governments differs widely but is relevant for obtaining adequate hardware, software and IT knowledge (structure). (4) Funds might stimulate municipalities with a small capacity, but in the past, Flemish funds were sometimes unilaterally abandoned, creating local distrust (history). (5) The configuration of Flemish departments stimulates silo creation: every department has its own budget. Information sharing challenges the classic revenue model (structure). (6) Flemish politicians see it as a means for cost reduction but *"a G2G IS is a current account. If there is a change in the IS, stakeholders have to invest money to adapt to these changes."* A lack of invested resources leads to suboptimal solutions. Ministerial priorities can influence the annual budget of an IS too (politics).

Support. (1–2) Funds to stimulate use are not desirable in times of budget cuts but may convince stakeholders to support an IS that has to outgrow technical problems. G2G ISs require immediate investment costs, benefits are only obtained over time. Funds are useful to bridge the period when other benefits cannot yet be reaped. (3) Cities have more means and a stronger bargaining power than small sized municipalities. (4) Support is given more easily when increased performance is expected. (5–6) If a ministerial cabinet supports an IS, it is easier to ask for more resources/cooperation of other departments. Respondents advise to prevent regular IS changes and to explain why these changes are needed.

Problem 2: Personnel. *Problem Description.* Every participating organisation in a G2G IS project should invest a sufficient amount of personnel time at all stages of the project, in order to tackle interoperability or IT problems and in order to enable data input and analysis.

Context. (1) In a G2G IS project the number of data inputters differs widely per organisation. 75 % of the Flemish municipalities has less than 20.000 citizens and takes up many tasks with few people. Cooperation in G2G ISs is cost demanding for small municipalities, resulting in higher investments than gains(structure). "*They chose their own priorities: not all data requests from the Flemish government will be answered."* (2) Governments often lack IT skilled personnel (structure). (3) E-government is demanding as it often requires a duplication of services (e.g. due to the digital divide or during the transition to an IS) (human factors and technological process). (4) Every governmental organisation works and evolves on its own speed (history).

Support. (1) Due to their size, cities have more negotiation power than smaller municipalities. Their ISs may be more advanced than the Flemish ones. Cities claim that if the Flemish government wants their support, it should recognise their expertise instead of imposing ISs. The respondents think that users with a small capacity can be motivated to support an IS by the availability of a help desk. (2–4) Difference in capacity/speed might be tackled by the creation of several entrance levels.

Problem 3: Intangible Resources. *Problem Description.* G2G ISs need to be fed with data from different parties.

Context. (1) Stakeholders in G2G context often face a lack of shared goals (environment). (2) The Flemish government aims to collect policy info for the whole of Flanders. The required data are available at local level but municipalities do not intend to spend their limited resources on making their data available to the Flemish government when their benefit is uncertain (environment). (3) This problem is exacerbated by previous experiences that it is hard to obtain data in return from the Flemish level (environment). (4) The Flemish government recently agreed to standardize on 'open data', but not all departments welcome this strategy. Some ISs are financed by 'pay per data use', but who will pay for 'open' ISs is not yet clear (environment).

Support. (2–3) The respondents notice that if the Flemish government wants local governments to support an IS, it should see them as data sharing partners and not as data subordinates. (4) Respondents think that stakeholder support for opening ISs is only realistic if politicians provide an alternative finance model for 'pay per use'.

4.2 Legal Agreements

Rules and legislation are mostly created to ensure quality, equity or responsible resource use. But these can restrict innovation if they cannot be easily adapted to specific needs or a dynamic environment [17]. In our research we found 3 recurrent legal problems that concern change in legislation, voluntariness of use and privacy issues.

Problem 4: Change in Legislation and Regulations. *Problem Description.* Legal or jurisdictional aspects may hinder the progress of e-government. Legislation and regulations need to be altered or sometimes developed.

Context. (1) The legal status of a digital G2G IS might not be recognised, even if it is, compared to its paper counterpart, the authentic source (structure). (2) Digitisation often comes after legislation. IS stakeholders may find their activities prohibited by formal rules. The option to simplify legislation is regularly ignored as idea champions lack time or the juridical capacity to perform such an exercise. As a consequence digitisation does not simplify the multiple adapted/expanded rules but builds yet another level of legislation (structure and environment). (3) Politicians change legislation/regulations regularly, which brings along an adaption cost for ISs (environment).

Support. (1–2) The respondents are less likely to support complex ISs. As such they believe that reengineering legislation is necessary. *"IT'ers tend to see legislation as hard to programme, holy and untouchable. They should realise it can be adapted."*

Problem 5: Voluntariness of Use. *Problem Description.* An IS needs to be used to be successful. Voluntariness of use differs per case, rights and obligations may be laid down in rules and regulations.

Context. (1) The Flemish government often legally obliges IS use for local governments. Non-use is not always reprimanded (regulators). The "bell tower principle" states that if the Flemish government asks for municipal efforts, it should provide a financial reward, this is often ignored in practice (history). (2) Another option is using the carrot by funding data exchange, yet this created perverse effects in the past (e.g. minimal data import to get the money) (environment). (3) A third option is voluntary participation in an IS based on a win-win (environment). (4) The need of local data for the Flemish government often originates from European requests (regulators).

Support. (1) Due to a lack of resources many local governments will not provide support to low priority G2G ISs. Use of legislation alone is an indolent solution for getting support, it may result in ISs plagued with poor data quality. As local governments become more emancipated, they tend to refuse support even if it is legally obliged. (2) Support can drop severely when funding stops. (3) Municipalities are tired of double data requests, they want to support data reuse. The respondents also advice to ask potential users who refuse to cooperate why this is the case (Fig. 1).

CONTEXT	
HUMAN FACTORS	TECHNICAL PROCESS
- Digital divide: duplication of services induces duplication of data input efforts	- Duplication of services induces duplication of development efforts
HISTORY	POLITICS
- Unilaterally adapted funds, low local trust - Organisations work at their own speed - Ignorance 'bell tower principle'	- Politicians underestimate (recurrent) costs - Ministerial priorities influence finances - Changes legislation ⇔ adaption cost ISs
ENVIRONMENT	STRUCTURE
- Eco. crisis, dropped willingness to do sth. for others - Culture: hard to locally obtain Flemish data - Interests: distrust, lack of shared goals, desire for benefits in return for data ⇔ obliged use - Interests: win/ funds as stimulus ⇔ perverse effects - Interests: Vagueness privacy as an excuse for not sharing data, open data ⇔ pay for use - Regulators: no reengineering, data sharing may be prohibited or obliged. Problem legal status digital ISs - Regulators: European data requests	- Differing budgetary & IT capacity - Budgetary configuration stimulates silos - Low (IT) personnel capacity 75% municipalities, own local priorities - Time intensive to change legislation, lack of juridical personnel capacity - Time consuming to get permission of (conflicting) privacy commissions

Fig. 1. Overview of legal and economic context constraints of Flemish G2G IS projects

Problem 6: Privacy. *Problem Description.* Governments that share information have to respect the fundamental right of privacy.

Context. (1) In order to share information, G2G IS projects often need a permission of a privacy commission, but obtaining permissions is time consuming. The more because the Belgian federal structure may make several privacy commissions competent according to the applicable policy level or policy domain. Respondents complain that the advice of different commissions may conflict (structure). (2) Data may not be shared or reused due to data protection regulations. It is still rather unclear if reused data may be enriched. The respondents see privacy as a vague issue (environment).

Support. (1–2) An IS often needs the support of a kind of privacy commission(s), which can be time consuming. The respondents ask for more clarity on privacy issues. They propose to only let the most relevant commission decide, when several privacy commissions are authorised. The respondents believe that on the other hand, privacy protection may also be a misused argument to refuse support to provide data.

5 Discussion

3 economic and 3 legal recurrent problems for Flemish G2G IS projects were uncovered by the analysis. Economic agreements have to be made to prevent problems with (in)tangible resources such as (1) money, (2) personnel, (3) data/information. Agreements on legal/juridical issues are crucial to prevent problems with: (4) too much/few changes in legislation/regulation, (5) voluntariness of use and (6) privacy. Several researchers confirm the importance of these problems: In terms of the economic perspective, previous research shows that it is hard to decide who will bear the costs of G2G IS projects [13], and that a lack of capacity is indeed a major barrier [9, 16, 18]. Financial support may promote ISs but obtaining funds can be tough [13]. Intangible resources are valuable, stakeholders want a benefit in return [1]. Concerning the juridical perspective, other scholars also found that G2G ISs may mismatch formal rules [11, 15] and require a time-intensive creation/adaption of legislation/regulations [11, 19]. They confirm that rules may be used as an excuse to block projects [17]. While ISs based on voluntary collaboration have more chance to succeed. Finally, the lack of clarity in privacy policies is a barrier for G2G ISs [9, 25].

This article investigated recurrent economic and juridical problems of G2G ISs in Flanders. It is an addition to a previous article about recurrent technological and political problems. Taking a look at the whole picture is interesting. All four categories of problems are based on the study of interviews in 2012 and focus groups in 2014. In total we found 14 recurrent problems. Each problem on itself is not new, factor research has listed these as well [21]. Yet, we go further by describing influencing context and support issues of these problems. By conducting a needs and support-power analysis it became clear that seemingly controllable problems have much deeper roots. Even more, context and support elements of different problems are interrelated.

In the previous paper elements pointed to: (1) a tendency of Flemish idea champions to merely focus on Flemish interests, (2) a lack of coordination of IS initiatives and (3) a political disinterest in ISs. These discourage local stakeholders to support Flemish G2G ISs. This paper studied if a needs and support-power analysis of economic and

juridical recurrent problems can provide additional insights: First, the previous findings are confirmed, the 4 categories of problems seem to have similar roots: (1) The ignorance of the bell tower principle, a unilateral adaption of funds and a difficulty to locally obtain Flemish data, point to the tendency to merely focus on Flemish interests. (2) A budgetary configuration that stimulates siloisation and a lack of (G2G) shared goals and benefits, point to a lack of integration. (3) The unrealistic expectation of quick savings without much expenditures points to a political disinterest and lack of knowledge of ISs. Second, the analysis of economic and juridical factors adds two root elements. Local stakeholders seem also discouraged to support Flemish G2G IS projects because of (4) major differences in organisational capacity (e.g. differences in budget, IT, personnel capacity or e-government speed) and (5) juridical complications such as complex piled legislation, the need for duplication and the insecure legal status of ISs. Flemish idea champions may become discouraged by conflicting privacy commissions and by the prohibition of data sharing. Knowing why local stakeholder support may be discouraged is a start, but it is essential to know which action Flemish idea champions can take to tackle recurrent problems at their roots:

Action 1: Move from Hierarchical Project Management to Network Process Management. The main focus on the own Flemish interests and a lack of coordination be-tween different departments point to a 'hierarchical-project' way of managing. Project managers focus on clear goals and a predefined output while minimizing the risk of random events and ignoring the process [7]. This approach is not compatible with the environment of G2G IS project in Flanders. A 'network-process' management approach would allow to actively involve local stakeholders, as such Flemish idea champions could get an image of their core values. They may also get more conscious of the world around them and might better coordinate actions which allows better conflict anticipation. Local and Flemish stakeholders continually interact and adapt to their environment. These dynamics make G2G IS projects difficult to manage. Managers must also realise that the involvement of many people risks scope creep [12].

Action 2: Go for the Win-Win: Manage Needs, Capacities and Speeds. Often legislation mandates without providing resources, which negatively influences mutual trust [11]. We notice a trend of municipalities to refuse cooperation even if it is legally obliged. They argue they are busy with their own business and have other needs [13]. Flemish idea champions are dependent from other parties. Our research results show that local and Flemish stakeholders expect some benefits in exchange of their data (e.g. funds, increased performance, less duplication or data reuse). If funds or a legal obligation are the only drivers to provide data, perverse effects (e.g. poor data quality) might surface. A win-win between partners is definitively more stimulating [15]. Bigdeli et al [4] state that the resources of local governments are more limited than those of central/regional governments. Yet even within one group of stakeholders, in this case municipalities, there is a large capacity difference: the smallest Flemish municipality counts 80 inhabitants, the largest 480.000. Populous local governments, with many resources, are more likely to adopt e-government than smaller less resourceful municipalities [25]. In general, it is hard for Flemish idea champions to deal with the variety of needs, capacities

and speeds [10]. This might be overcome by involving small, medium and large local governments and providing several IS entrance levels.

Action 3: Make Politicians More Aware of their Juridical Deeds and the Interaction with ISs, Dare to Redesign Legislation. Bekkers believes that the design of an IS should come before the creation of new legislation [1]. Flemish politicians are continually generating new legislation, which influences organisational flexibility [4]. If there is a change in legislation, business processes and their supportive systems have to be adapted [19]. G2G IS projects are burdened by a web of detailed and even conflicting rules/regulations [1]. The short term orientation of Flemish politicians and the bureaucratic nature of public organisations (e.g. focus on predictability, legal security and equality) frustrates innovation [10]. Flemish idea champions should try to make politicians aware of their ignorance of ISs and the related juridical complications.

6 Conclusion and Future Research

This paper studied the roots of six recurrent economic and juridical problems of Flemish G2G IS. We aimed to extend the body of knowledge by investigating how local contingencies and support-power relations affect the likelihood of failure of Flemish G2G IS projects. The research findings show that a needs and support power analysis (i.e. a process perspective) provides additional insights for G2G IS management in Flanders. The analysis takes a broader look on IS failure than the classic factor perspective by incorporating the dynamic IS context and its stakeholders.

A previous study of recurrent technological and political problems of Flemish G2G IS, showed that apparently controllable problems have deeper roots which discourage local stakeholders to support ISs (i.e. the focus on Flemish interests by idea champions, political disinterest in technology and a lack of coordination). In this study it appears that economic and juridical problems have the same roots. Yet these are also rooted in major differences in organisational capacities and juridical complications.

Insight in the deeper roots of recurrent IS problems adds, compared to factor research, a new piece to the complex IS failure puzzle. Given this knowledge, which action could managers take? Managers should consider that the 'network process' management approach might be more suitable than an 'hierarchical project' management approach concerning Flemish G2G IS projects, they should have an eye for needs/capacity/speeds and could make politicians more aware of juridical consequences of their deeds, and in particular for the interplay of juridical aspects with ISs.

Our research findings are limited to Flemish G2G IS projects. Future research could study the roots of recurrent problems of G2G ISs in other regions or countries.

References

1. Bekkers, V.: The three faces of e-government: innovation, interaction and governance. In: Anttiroiko, A-V., et al. (eds). Innovations in Public Governance. Series Innovation in the Public Sector, pp. 194–216. IOS Press, Amsterdam (2011)

2. Bekkers, V., van Duivenboden, H., Thaens, M.: ICT and Public Innovation. Assessing the ICT-Driven Modernization of Public Administration. IOS Press, Amsterdam (2006)
3. Belanger, F., Tech, V.: Theorizing in Information Systems Research: using focus groups. Australas. J. Inf. Syst. **17**(2), 109–135 (2012)
4. Bigdeli, Z., Kamel, M., deCesare, S.: Inter-organisational electronic information sharing in local G2G settings: a socio-technical issue. In: European Conference on ISs, paper 79 (2011)
5. Dawes, S.: E-government and innovation: a fresh look at experience. In: IFIP EGOV (2014)
6. Dawes, S., Cresswel, A.M., Pardo, T.A.: From ''need to know'' to ''need to share'': tangled problems, information boundaries, and the building of public sector knowledge networks. Public Adm. Rev. **69**(3), 392–402 (2009)
7. De Bruyn, H., ten Heuvelhof, E., in 't Veld, R.: Process Management, Why Project Management Fails in Complex Decision Making. Springer, Heidelberg (2010)
8. Dwivedi, Y.K., Wastell, D., Laumer, S., Zinner, H., et al.: Research on IS failures and success: status update and future directions. Inf. Syst. Front. **17**, 143–157 (2014)
9. Ebrahim, Z., Irani, Z.: E-government adoption: architecture and barriers. Bus. Process Manage. J. **11**(5), 589–611 (2005)
10. Edmiston, K.L.: State and local e-government. Prospects and challenges. Am. Rev. Public Adm. **33**(1), 20–45 (2003)
11. Fountain, J.E.: Challenges to organizational change: multi-level integrated information structures (MIIS). In: National Center for Digital Government, Paper 15 (2006)
12. Janssen, M., van der Voort, H., van Veenstra, A.F.: Failure of large transformation projects from the viewpoint of complex adaptive systems: management principles for dealing with project dynamics. Inf. Syst. Frontiers **17**(1), 15–29 (2014)
13. Jing, F., Pengzhu, Z.A.: Field study of G2G information sharing in Chinese context based on the layered behavioral model. In: HICCS 42, pp. 1–13 (2009)
14. Krueger, R.: Focus Groups: A Practical Guide for Applied Research. Sage Publications, Thousand Oaks (1994)
15. Lam, W.: Barriers to e-government integration. J. Enterp. Inf. Manage. **18**(5), 511–530 (2005)
16. Moon, M.J.: The evolution of e-government among municipalities; rhetoric or reality? Public Adm. Rev. **62**(4), 424–433 (2002)
17. OECD: Innovating the public sector: from ideas to impact, Paris, 1–22 (2014)
18. Pardo, T.A., Nam, T., Burke, G.B.: E-government interoperability: interaction of policy, management and technology. Soc. Sci. Comput. Rev. **30**, 1–17 (2011)
19. Pardo, T.A., Scholl, H.J.: Walking atop the cliffs: avoiding failure and reducing risk in large scale e-government projects. In: HICCS 35 (2002)
20. Peters, R.M., Janssen, M., van Engers, T.M.: Measuring e-Gov impact existing practices and shortcomings. In: 6th International Conference of ECommerce, pp. 480–489 (2004)
21. Petter, S., DeLone, W., McLean, E.R.: Information Systems Success: The Quest for the Independent Variable. Journal of MIS **29**(4), 7–61 (2013)
22. Sauer, C.: Why IS Fail: A Case Study Approach. Alfred Waller Publishers, Suffolk (1993)
23. Van Cauter, L., Snoeck, M., Crompvoets, J.: Removing the blinkers: what a process view learns about G2G information systems in Flanders. In: IFIP WG 8.5., Thessaloniki (2015)
24. Yang, T.M., Pardo, T., Wu, Y.J.: How is information shared across the boundaries of government agencies? An e-Government case study. Gov. Inf. Q. **31**, 637–652 (2014)
25. Yang, T.M., Zheng, L., Pardo, T.: The boundaries of information sharing and integration: case study of Taiwan e-Government. Gov. Inf. Q. **29**, S51–S60 (2012)

A Visual Uptake on the Digital Divide

Farooq Mubarak(✉) and Reima Suomi

Turku School of Economics, Department of Information Systems Science,
University of Turku, Turku, Finland
{farmub, reima.suomi}@utu.fi

Abstract. Factors found to influence the adoption of ICT have been explored in several studies. However, few writers have been able to produce any systematic research into the digital divide. Although, differences of opinion still exist, a growing body of literature has established that income and education are positively related to digitalization patterns. This research attempts to deepen the understanding of the present ambiguous relationship between socio-economic indicators and the ICT. This account tested the links between socio-economic variables (GDP per capita, GINI index, World Bank Education Statistics, and Transparency International's corruption perception index) and ICT diffusion across developed and developing countries. Positive correlations were found for income and ICT, education and ICT. A negative correlation was found between corruption and ICT adoption. The paper discusses implication of these findings and suggests future courses of actions for policy makers. Proceeding from the findings of this paper, this research suggests there is an urgent need to address the digital divide by initiating impactful efforts to reduce it.

Keywords: Digital divide · ICT · Digital technologies

1 Introduction

Researchers have long sought to determine how patterns of digitalization link with economic variables. To that end, a hefty volume of publications has appeared in academia that looks at such links that have a variety of well-disputed varying research methods and measurement mechanisms. Moreover, recent advancements in ICT research have laid rest to several myths concerning the nature of the digital divide. Nevertheless, much doubt still persists since there still appears to be little agreement on the leading causes of the digital divide. The past decade, in particular, has seen rapid progress in ICT research, which has reshaped the conventional concept of the digital divide from narrow to wide and made it considerably complex paradigm. This shift has resulted in mounting literature across academia, politics, and the press. Much of the existing debate on the digital divide revolves around the qualitative nature of the issue, while the quantitative uptake on the topic appears to be overlooked. Thus, Noh and Yoo [1] view the measurement of the digital divide as a controversial issue. Owing to the complexities of the subject, some researchers [2–4] state that the digital divide is a confused theme in literature. These critics justify their views by noting severe pitfalls of the data and methodologies used to quantify the digital divide.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 398–415, 2015.
DOI: 10.1007/978-3-319-25013-7_32

In the history of the digital divide, poverty has been thought of as a key factor that is responsible for the breach in the access and the utilization of the digital technologies. In their major analysis of publications from thirty leading researchers, Skok and Ryder [5] concluded that GDP per capita and education were the principal factors responsible for the digital divide. A similar result was reported by Billon et al. [6] who found GDP and infrastructure as to be the main factors for slow digital progress in developing countries. A number of researchers [e.g. 7–9] have acknowledged GDP per capita/income as a leading cause of the digital divide between and within countries. It appears safe to say that poverty alone explains a major portion of the digital divide outgrowth.

The connections between GDP per capita and patterns of digitalization have been an object of research since the evolution of the digital divide concept, in the mid-1990s. Vodoz et al. [10] noted that individuals with higher education levels are likely to adopt the digital technologies faster than people with low or no education at all. Two large-scale studies [11, 12] demonstrated the positive correlation between GDP and the Internet diffusion curve.

However, uncertainty still exists about the relation between education and ICT diffusion; and two major studies [13, 14] defy any relation between the two. Consonantly, Stanley [15] intensifies psychosocial resistances as key factor responsible for the digital divide, thus putting aside income and education. Nevertheless, considerable criticism has been levelled at quantitative research on the digital divide. Much of the research on this subject has been restricted to local and limited comparisons, while only a few studies such as those by Cruz-Jesus et al. [14] have attempted to assess the digital divide across a wide range of geographical territories.

The recent rise in public computer facilities in developed countries has allowed a large segment of the population to benefit from ICT, who otherwise could not afford computers and the Internet. Thus, it can be fairly argued that there exists a category of people in developed countries who can benefit from ICT regardless of in-come constraints. If the findings made by Vodoz et al. [10] about the high rates of ICT adoption on education being dependent on education are accurate, then what explains the exponential growth of the Internet in developing regions with low literacy rates, such as South-East Asia? What is not yet clear is the measured impact of income and education on ICT adoption in the current era. Despite the research mentioned earlier, there is still very little scientific understanding of the degree of the relationship if any, of socio-economic indicators with ICT, particularly in developed countries. In addition, the general dispute in the quantitative literature on the digital divide hints that much of the evidence so far is inconclusive at best. This indicates a need to understand the various perceptions of any possible connections of socio-economic indicators with ICT in this current age. For the purposes of this study, the chosen socio-economic variables are the World Bank GDP per capita, the GINI index, World Bank Education, and Transparency Inter-national's Corruption Perception index.

Even though, some research has studied the socio-economic links with ICT, most research has been undertaken by analyzing just a few statistical observations. Nonetheless, it is possible to further improve the research design and scope by utilizing a combination of methods. With this goal, the present research seeks to obtain visualizations of the major factors that are responsible for the digital divide. Drawing upon

this stand of research into the digital divide, this paper shall attempt to verify the aforementioned claims in preceding paragraphs by examining the links between income, education, and corruption with ICT patterns across high-income Nordic countries, the low-income Indian sub-continent region, and a few middle-income countries. A secondary aim is to shed light on the implication of the findings and suggest a direction for possible future developments. This paper aims to broaden the scope of the main factors responsible for the digital divide, since the digital disparities are intermixed with social and psychological factors in addition to income and education. The hypothetical premises at this point rest on four assumptions: GDP per capita is positive related to Internet adoption, GINI index is inversely related to Internet adoption, education is positively related to Internet adoption, and Corruption is inversely related to Internet adoption.

2 Literature Review

During the past twenty years, much more information has become available on the digital divide and its rigorous threats to the world economy. However, very little was found in the literature on the question of quantitative analysis of the digital divide. The digital divide can be defined as the disparity between those who do and do not benefit from digital technologies [16]. In the context of the digital divide, the chief division is between two significant groups namely inclusion and exclusion with respect to those who benefit and those who do not benefit from digital technologies. One group consisting of developed countries is continuously reaping the benefits of ICT whereas the other group consisting of developing countries is missing out on many benefits due to lack of access to the digital technologies [17, 18]. Menou [19] and Mansell [20] already warned that if the issue of the digital divide is left unattended, world inhabitants will be living in a dual planet. Call for future research on the nature of the digital divide has been a recurring theme of many scholarly articles.

As noted in the introduction that digital divide has been viewed as a confused theme in literature: some studies [19, 21, 22] consider the digital divide to be a matter of gap in access to ICT while others [1, 23–30] consider the digital divide as a complex and broad phenomenon where several variables play their respective part. Pick and Nishida [31] found education to be the principal determinant of technology utilization. It can be inferred from the study´s conclusion that the role of education is significant in increasing adoption of digital technologies. Similarly, Cooke and Greenwood [32] maintain that the educational sector has made significant progress in promoting the adoption of ICT. Pittman [33] postulates that the role of ICT is essential in fostering a globally diverse educational system.

However, recently some literature has emerged that offers contradictory findings regarding the role of education in ICT adoption. An empirical investigation led by Lee [34] suggests that demographic factors (age, gender, education) have little effect on the digital divide. Unlike Pittman [33], a survey study of 158 Small and Medium Enterprise (SME) owners by Middleton and Chambers [13] found that education has no effect on the adoption of Internet. However, this attitude would appear to be outdated even the study was published in the year 2010. There are limits to how far the idea of

Middleton and Chambers [13] can be taken because ICT is being increasingly incorporated into the education systems worldwide.

Tipton [7] and Olaniran and Agnello [8] document income disparity as the leading cause of the digital divide by noting that the digital divide reflects high income levels in developed world where as the opposite is true for the developing world. Quibria et al. [12] report a strong correlation between GDP per capita and the usage of computer. However, Tavani [27] develops the claim that there are numerous other factors responsible for the breach in access and utilization of ICT other than income alone. Brooks et al. [9] maintain that costs of Internet connectivity in developing countries is significantly higher than those in developed countries. This corroborates with the view of Norris [35] who maintain that the richer countries are better in reaping the benefits of ICT than the poorer countries. The evidence presented in this paragraph suggests that there is a strong connection between GDP per capita and the patterns of digitalization. However, a number of studies [e.g. 14, 36–39] have reported significant digital divide in developed countries, which questions the relation of GDP with the ICT diffusion. This doubt is further reinforced since some researchers [40, 41] report a regional digital divide with respect to urban and rural settlements in high-income developed countries.

Novo-Corti et al. [42] maintain that simply promoting the access to digital technologies is a simple solution to overcome the digital divide. Although, the study was targeted towards a particular region, the claim is question-able because the digital divide has been proved as a complex phenomenon and a variety of factors are responsible for the di-vide other than just access. Bach et al. [43] calls for effective policies for organizations and governments to com-bat the digital divide. This corroborates with the findings of Graham [44] who highlights the need of effective government systems with effective subsidies to minimize the digital divide. Peng [45] points out that although governments have access to household profile data such as education, income, and gender, they often lack reliable insights into psychological and cognitive profiles of individuals. This implies that efforts to fight the digital divide are required on multiple fronts since other than in-come and education; psychological factors are also responsible for adoption/non-adoption of ICT.

The insights drawn from the literature review advances us towards testing three hypothesis. First, income per capita bears a positive relation with ICT adoption. Second, education has a positive link with ICT adoption. Finally, corruption bears negative relation with ICT adoption.

3 Research Design

3.1 Methodology

To date various methods have been developed and introduced in order to determine the connection between two or more variables. Traditionally, the digital divide measurements have been studied by comparing the variables in a select geographical territory. In addition, correlations between these variables have also been calculated to measure the degree of relationship. In the present study, the visualization approach was chosen to represent the association of GDP and education with the digitalization patterns across OECD member countries. The correlation attempt to provide an estimate of the

degree of association between the variables under consideration. The correlations were calculated by using the following formula.

$$r = \frac{1}{n-1} \Sigma \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right)$$

Where r is the coefficient of correlation, X and Y are variables, whereas x bar and y bar represent respective sample means. S represents the sample standard deviation of the respective variables x and y. n equals the number of items in the samples under consideration. Pearson product correlations from the above mentioned formula were calculated between the variables under consideration after each figure. The average values of the data sets were taken for the available years, wherever possible. In some cases, data was missing for some countries in specific years; thus only years with complete data for all countries were taken into consideration.

3.2 Data Selection

Although, a range of different organizations provide data sets for GDP and similar economic indicators, this research employed data from World Bank statistics database. In addition to the simplicity of the data downloads according to the customized preferences, World Bank statistics are known for providing credible information. The statistics were also found to be in close accordance with data provided by other institutions, although there were minor differences. PISA test scores were taken directly from the official PISA scores website. The corruption Perception Index was taken from the Transparency International official website. For the purposes of the present study, information from the developed Nordic countries, developing Indian subcontinent countries and a few middle-income countries was chosen. In some instances, the data for some countries was not available for a particular indicator; therefore different countries were added to the comparison. The data sources are detailed in Table 1.

Table 1. Data sources

Name	Institution	URL	Selected data source	URL
Education efficiency	PISA-test scores	www.oecd.org/pisa	Scores in mathematics	http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm
Income	World Bank	www.worldbank.org	GDP per capita	data.worldbank.org/indicator/NY.GDP.PCAP.CD
Income inequality	World Bank	www.worldbank.org	GINI index	data.worldbank.org/indicator/SI.POV.GINI
Corruption	Transparency International	www.transparency.org/country	Corruption Perception index	www.transparency.org/research/cpi/overview

3.3 Visualizations

Human brains are designed to process visual images before texts, and they need less energy to consume images than texts [46]. In particular, visualizations present clear pictures of possible trends where the data is vast. Therefore, visualizations were chosen as a means for seeing any connections between the variables. In addition to giving a clear picture of a major scenario, visualizations can also represent the predictability of a certain variable's behavior over time.

In the present study, there are a mix of developed and developing countries under consideration. After trying various visualizations in Microsoft Excel software, a few were selected because they tend to report the best trends between the variables. Charts were used to show any possible link between the two variables.

4 Results and Discussion

4.1 Results

The visualizations revealed several interesting insights, a few of which negate previous studies. In the forthcoming paragraphs, each figure shall be discussed and all possible interpretations shall be subsequently drawn in the next section of discussion (Table 2).

Figure 1 shows that a positive parallel relation was found between GDP per capita and Internet users per hundred people in developed and developing countries. The Pearson product correlation coefficient "r" was found to be 0.882. The positive correlation suggests that there is a tendency for increase in the number of Internet users with an increase in the GDP per capita of a given country. It can be seen from figure that Norway has the highest GDP per capita and the highest percentage of Internet users. Along with Norway, other Nordic countries show high rates of GDP per capita and percentage of Internet users. This is in stark contrast with the Indian subcontinent countries which show low rates of GDP per capita and Internet users (Table 3).

Figure 2 compares the GINI index with the Internet user's percentage across a range of developed, developing, and middle-income countries. A general trend that can be noticed from the Fig. 2 that higher the GINI index, lower the Internet users percentage. For instance, Iceland has the lowest value of GINI index and the highest number of Internet users percentage along with Norway. Thus, it can be inferred that higher the inequality of income in a given country, lower shall be the ICT adoption, in general. There are however exceptions to this rule. For example, GINI index values of Nepal and Bangladesh are almost same, however Nepal shows considerable high percentage of Internet users. This negative result might be due to other factors such as political and regional infrastructure differences. The correlation analysis yielded a negative value of −0,45 indicating a negative link between inequality of income and Internet usage.

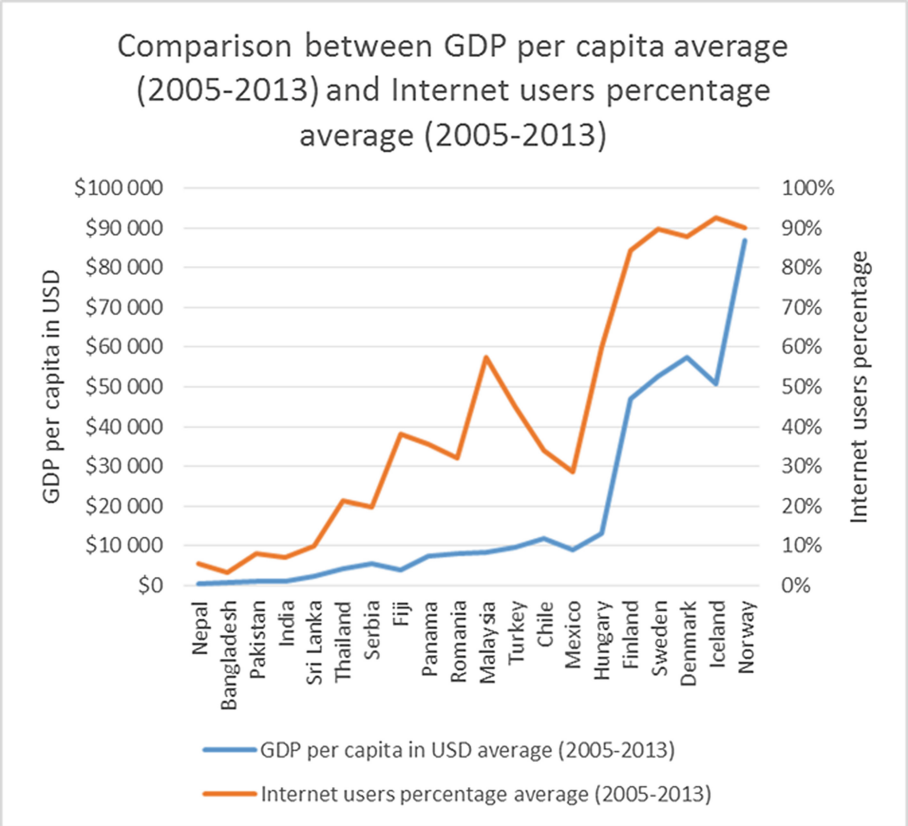


Fig. 1. Comparison between GDP per capita average (2005–2013) and Internet users' percentage average (2005–2013) (Based on World Bank 2015)

Table 2. Correlation between PISA 2012 test scores and the percentage of Internet users

Emerging correlation in Fig. 2	
Correlation between GDP per capita average (2005–2013) and Internet users percentage average (2005–2013)	0.882

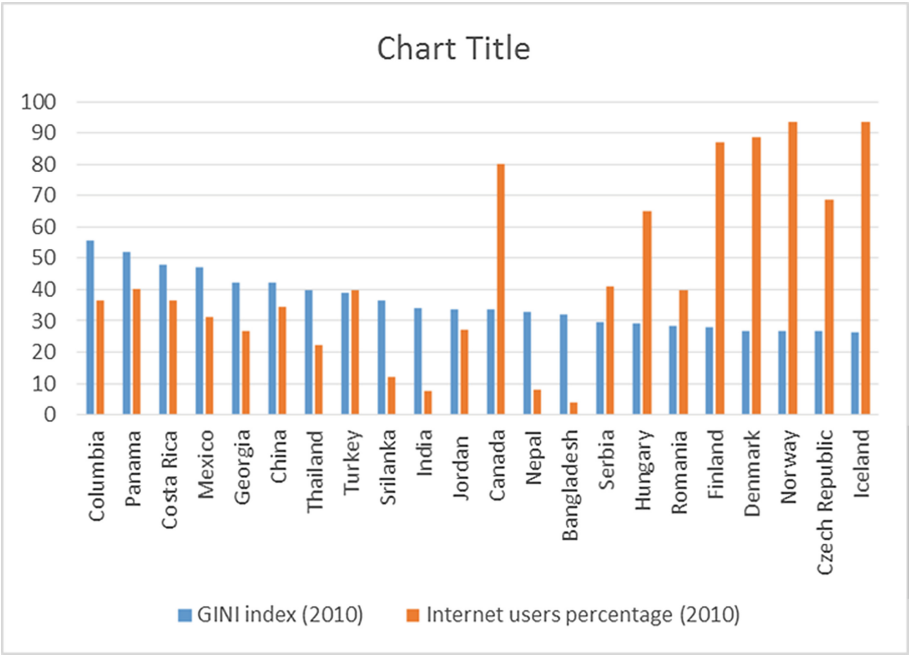


Fig. 2. Comparison between GINI index 2010 and Internet users 2010 (Based on World Bank 2015)

Table 3. Correlation between GINI index in percentage (2010) and Internet users in percentages (2010)

Emerging correlation in Fig. 2	
Correlation between the GINI index in percentages (2010) and Internet users in percentages (2010)	−0.45

Figure 3 compares percentages of secondary school enrollment and Internet users across Nordic, few Indian subcontinent countries for which data was available, and middle-income countries. Comparatively less intense than the correlation between GDP per capita and Internet users, however, still a positive correlation between secondary school enrollment and Internet diffusion was found to be 0.699 as reported in Table 4. All Nordic countries show high percentages of secondary school enrollments and

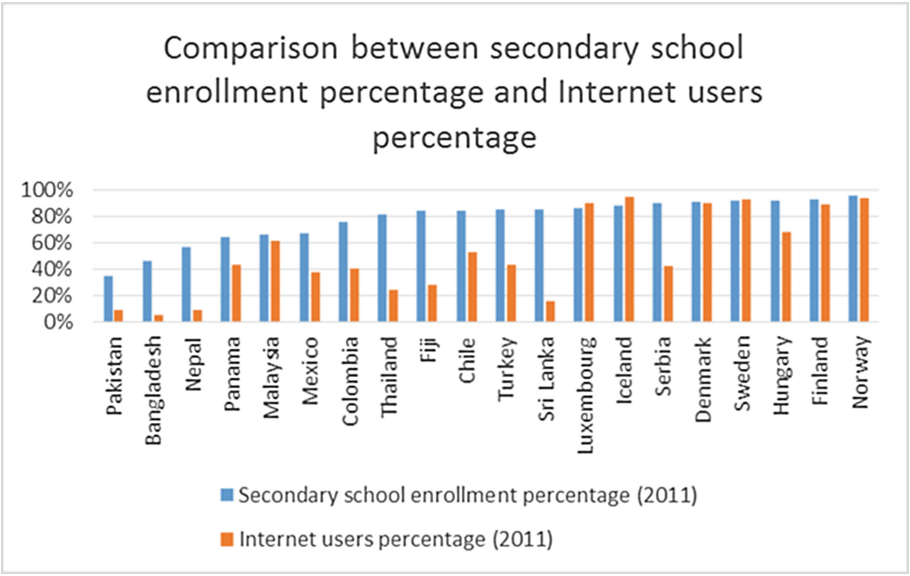


Fig. 3. Comparison between the percentage of secondary school enrollment and the percentage of Internet users for the year 2011

Table 4. Correlation between the percentages of secondary school enrollment and the percentage of Internet users

Emerging correlation in Fig. 3	
Correlation between the percentages of secondary school enrollment and Internet users	0.699

Internet users, whereas countries in Indian subcontinent region show low values of both variables under examination. There are however exceptions to this general trend of parallel growth of secondary school enrollments and Internet users. For instance, Luxembourg and Sri Lanka have almost similar values of secondary school enrollments at 86,4 % and 85,4 % respectively but they differ remarkably in percentage of Internet users. Despite the high secondary school enrollments rate, Sri Lanka has only 15 % of Internet users as compared to 90,02 % in Luxembourg.

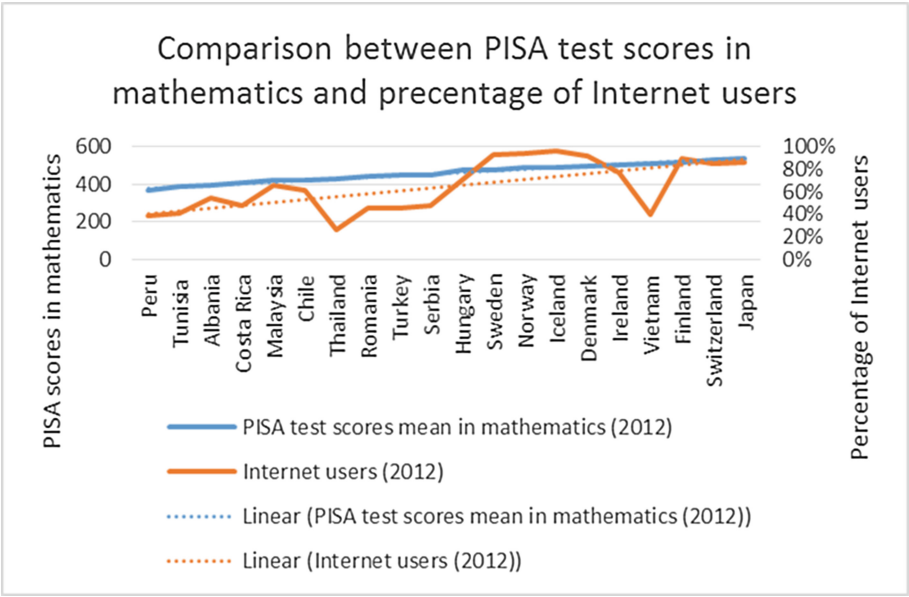


Fig. 4. Comparison between PISA 2012 test scores in mathematics and the percentage of Internet users (Based on data from OECD 2015 and World Bank 2012

Table 5. Correlation between PISA 2012 test scores and the percentage of Internet users

Emerging correlation in Fig. 4	
Correlation between PISA 2012 test scores and Internet users percentage	0.693

Advancing the comparison between education and ICT, PISA test results were included in the analysis. Since, the Indian subcontinent countries did not take part in PISA test, a different set of developing and developed countries were taken into account. It can be inferred from the Fig. 4 that countries with high PISA test scores have high percentage of Internet users, suggesting a positive relation between the two. However, there are exceptions to this assumption. For example, Vietnam and Finland show almost same values of PISA test scores, however they vary considerably in terms of Internet users percentage. Finland has far higher percentage of Internet users at 89,88 % than Vietnam at 39,50 %. The correlation between the subject variables in Fig. 4 was found to be 0,693 as reported in Table 5. The positive value of the correlation at 0,693 suggests a positive link between PISA test scores and Internet usage. The findings from Figs. 3 and 4 are close to the previous findings of Vodoz et al. [10] and Skok and Ryder [5] who determined education to be a principal factor responsible for ICT adoption.

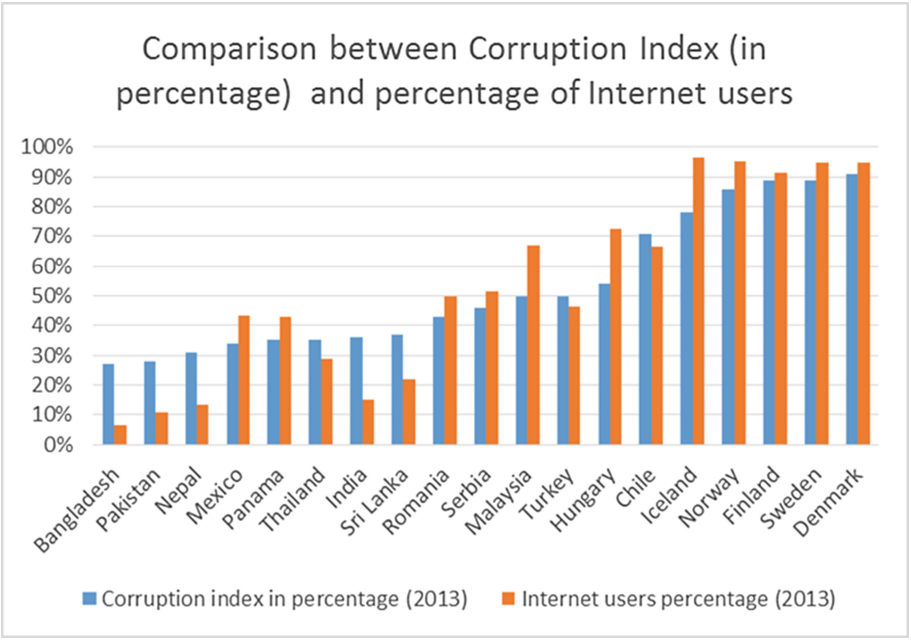


Fig. 5. Comparison between Corruption perception index and Internet users (Based on data from Transparency International Secretariat 2015 and World Bank 2015)

Table 6. Correlation between Corruption perception index and Internet users

Emerging correlation in Fig. 5	
Correlation between Corruption perception index in percentage (2013) and Internet users (2013)	0,933

Figure 5 illustrates how the corruption affects the percentage of Internet users. Corruption perception index shows the percentage of cleanliness in a given country. Figure 5 depicts that Denmark is the cleanest country in terms of corruption and has a high rate of Internet users. This is closely followed by Finland, Sweden, and Norway who are clean from corruption and show high Internet users percentage. On contrary, Bangladesh is the most corrupt country and has the lowest percentage of Internet users. It is apparent from this figure that more the cleanliness from corruption in a given country, more the percentage of Internet users. Correlation between Corruption perception index and Internet users was found to be 0,933. The high positive value of correlation at 0,933 indicates a significantly strong correlation between cleanliness from corruption in a country and Internet usage.

Taken together, these visualizations and correlations lend some support to test the hypothetical assumptions introduced in the beginning of the study. The results of the

hypothetical testing is tabulated in the Table 6. In summary, the results support the assumptions presented in the introduction. The next section, therefore, moves on to discuss the findings and their implications in detail (Table 7).

Table 7. Results of hypothesis

Hypothesis	Measuring factor	Hypothetical statements	Calculated correlations	Results
H1:	Income	GDP per capita and the rate of Internet use are positively related	0.89	Supported
		Inequality of income (GINI index) is inversely related to Internet user	−0.45	Supported
H2	Education	Secondary school enrollment and the rate of Internet use are positively related	0.699	Supported
		PISA test scores and the rate of Internet use are positively related	0.693	Supported
H3	Corruption	Corruption is inversely related to the rate of Internet use	0.933	Supported

4.2 Discussion

This study set out to determine the possible links between socioeconomic indicators and ICT in developed and developing countries. It was hypothesized that income and education are positively related to ICT adoption whereas inequality of income and corruption were negatively associated with ICT adoption. As mentioned in the introduction and literature review, the abundant studies on the digital divide show no definite consensus on above mentioned links to draw any firm conclusions. Results of this study indicate that indeed income and education have a positive link with ICT whereas corruption bears a negative relation.

As shown in the literature review, there has been an inconclusive debate about whether education bears any considerable link with ICT or not. Although, there have been a few dissenters to the view that education and ICT bears any significant relation; the present research dispels this view by noticing a positive correlation between education and Internet use from two stand points. First, secondary school enrollment appears to be positively related to Internet use. Second, countries with high scores of PISA test results often show high penetration of Internet.

Correlation between GDP per capita and Internet diffusion was significantly higher than the correlation between secondary school enrollment and Internet diffusion. This leaves room for interpreting that income disparity is the ruling factor responsible for the digital divide. The result emerging from the correlation between educational factor and Internet diffusion is consistent with the views of Le (2010, 84), who maintains that education has little effect on the adoption of digital technologies. However, reader must

bear in mind the ease of access to public free Internet in developed countries; opposite results are likely for developing countries where one may only/mostly tie knot with digital technologies for the purposes of higher education.

A possible explanation for a slight degree of correlation be-tween education and ICT might be that modern modes of education themselves encourage the adoption of digital technologies. ICT has been integrated with education on a global scale excluding misfortunate poverty ridden areas. It then follows that in order to continue the education; one has to utilize ICT, which may be the main motivator behind the purchase decision of digital technologies. This finding has important implications for strengthening the education systems particularly in the developing countries. Education and ICT in several developing countries are ridiculously expensive leaving millions of masses behind, for them needs of education and ICT are pushed aside by basic human needs of food and shelter. Poverty once again wins in breeding nuances of low standards of living including considerable portion of the digital divide pie. Governments particularly in the developing regions should therefore concentrate on providing ICT-enhanced education at reasonable costs for masses. Ideally, however, the solution should rest somewhere near providing free education and access to ICT wherever it is feasible.

The results of this study will now be compared to the findings of previous work. Present findings are consistent with earlier re-search (Tipton [7]; Norris [35]; Olaniran and Agnello [8]) which documents income disparity as the principal reason for the digital disparities in the world. While the current findings about positive relation between education and ICT corroborates with Lee [34], they negate with the results of Middleton and Chambers [13] who report that education has no effect on the ICT adoption. This difference urges itself as an evidence of dispute in the research concerning factors responsible for the digital divide.

Turning now to the hypothesis posed at the beginning of the study, it is now possible to state that income factor rules as the leading cause of the digital divide: however, education has slight effect on ICT adoption. Current research appears to validate the assumption that corruption and inequality of income are negatively associated with ICT use. This combination of findings provides some support for the conceptual premise that poverty is mainly responsible for breeding the nuances of the digital divide. An implication of this could be to provide subsidies on ICT related products so that masses can reach and bene-fit from the digital revolution.

Among various lasting divides on accounts of wealth, health and standards of living, the digital divide is becoming increasingly difficult to ignore due to strong ties of ICT with economic growth and wellbeing. Despite massive progress in the digital technologies over the past twenty years, the ICT diffusion re-mains regrettably uneven at a global scale. The underlying threats posed by the digital divide have been globally acknowledged across press and policy discourses, resulting in the production of various publications and action plans. Fortunately, there are good policies by OECD and International Telecommunications Union in the battle against the digital divide; unfortunately, the action plans in these policies are often poorly calibrated across different countries. Among various reasons, one obvious reason for inappropriate action plans by governments lies in the fact that the quantitative nature of the digital divide is significantly overlooked in the literature.

Researchers seem to have routinely confused technology access divide with the digital divide over the span of a decade since the evolution of the digital divide concept in the mid-1990s. The heated debate on the digital divide over the past decade left a considerable room for interpreting the digital divide as a broad and complex phenomenon. Soon it was established that breach in technology access was a small portion of the giant digital di-vide; the need for further research into the digital divide be-came even significant than earlier. Today, the digital divide breaks along multiple fronts ranging from individual to a global scale. This broadly indicates that the digital divide can be best tackled by initiating efforts from multiple ends.

As was noted in the introduction of this paper, poor quantitative understanding of the digital divide seriously impedes governments' abilities to form appropriate frameworks to minimize the digital divide; it becomes increasingly difficult to ignore the poorly understood quantitative aspect of the digital divide. Measurement of the digital divide is a classic problem often noted for deployment of unreliable data in the analysis. Once an adequate quantitative understanding of the subject is established and universally acknowledged, the mission to minimize the digital divide can be crystallized faster than ever before.

Prior research has stressed enough that governments should deploy superior strategy by focusing on weakening the roots of the digital divide rather than just providing access to ICT assuming that the market forces shall eliminate the digital divide over time. Unless governments execute impactful policies, the digital divide shall remain a dilemma for already troubled world economy.

5 Conclusion

5.1 Summary of Findings

This study was undertaken to verify the links of socio-economic indicators concerning GDP per capita, GINI index, education, and corruption with ICT. The present paper has given an account of validated positive links between income along with education and Internet usage. While there has been considerable research on the digital divide, only few studies have attempted to investigate the quantitative nature of the digital divide on a large scale. The doubt in existing quantitative accounts of the digital divide is reinforced by the distaste several researchers have in the data and methodologies used for analysis. Therefore, this study has taken a unique visualization approach with measured degree of covariance to provide a confirmatory evidence of any emerging relation between the factors under current examination.

Paired with literary clues, the visualizations paint a compelling picture in support of the hypothesis posed at the beginning of this study. This study has shown a positive association of income with levels of ICT penetration and a marginal correlation between education and ICT across an array of low-income, middle-income, and high-income countries. Present research has also shown that corruption has is inversely related to Internet usage. The present study confirms the previous findings about income and ICT relation and contributes additional evidence that the increasing GDP per capita determines increasing ICT adoption rate. Among the plausible explanation

for this finding is that high purchasing power encourages investment in general. Nevertheless, the results of this study do not support the idea that education and ICT are very highly related with each other. It is, however, not an inalterable rule because in some countries, poor can benefit from free public access to computers and Internet regardless of income/education levels.

The work contributes to existing knowledge in digital divide by strengthening the views of previous research and noting that there exists a dispute in literature regarding the relation of education and ICT. However, low degree of correlation between education and ICT found in this research cannot be extrapolated to the developing world where there is not widespread free public access to ICT and one must purchase the digital technologies for education purposes. The present findings also leave significant room for blaming poverty as the leading cause of the digital divide. It is possible to state that poverty alone breeds a significant proportion of the digital divide; nevertheless other factors such as motivation to adopt ICT, education, forced adoption of ICT due to work requirements, and cultural norms are important to consider when addressing the issue of the digital divide.

The practical implication of these findings is that governments must start efforts on multiple fronts to round up the economic threats posed by the digital divide. On the basis of the evidence currently available, it is possible to suggest that the education in developing countries should be made easily affordable to the poor masses if not totally free. While the poverty breeds the digital divide, the digital inclusion would breed economic development due to enhanced workforce. Overtime, the economic and social benefits of affordable education and ICT shall outweigh the sacrifice in monetary costs by the governments.

Press and policy documents in the name of effective policies against the digital divide shall remain fruitless without pronouncing a decisive aggression against the digital divide on a global scale. The relentless objections to the existing accounts on quantification of the digital divide make it advisable to reconsider the methodology and data used for analysis in future research on the topic. It appears that the crowning success in the battle against the digital divide would require a thorough grasp of both quantitative and qualitative nature of the digital divide. It is a high time to change the course of the digital divide history by initiating impactful efforts on multiple fronts, preferably on determining the mechanism to accurately quantify the digital divide. The digital divide has caused severe havoc to socio-economic lives of millions of people, not to mention its deadly impact on economic footings. Unless there are concrete moves by the governments against the digital divide, the gap between inclusion and exclusion groups shall continue to pose rigorous threats on the world economy.

5.2 Limitations

The findings in this paper are subject to at least two main limitations. First, the chosen data was only from World Bank database in addition to official PISA website and Transparency International website; a wiser approach could have been to perform the statistical analysis with data from different databases and then compare the results. Second, few variables representing education such as PISA results and secondary

school enrollments were included in analysis. The present research might have been enriched by including several representative variables of education and ICT diffusion in analysis.

5.3 Future Directions

A natural progression of this work would be to analyze the links tested in this study in the developing world. There is abundant room for future progress in determining the links between ICT and other indicators of the economy. Research questions that could be asked include link of ICT with education at different levels, effect of free public ICT access enters on the relation between income and ICT diffusion, comparing the education and income relations with ICT in OECD member states with developing countries. Future work on exploring such connections with different variables shall help to understand patterns of ICT diffusion from multiple perspectives. Further research might explore the links of cultural influences and cognitive factors with ICT adoption on a broad scale including developing and developed countries. The present slow progress in the digital divide projects a poor quantitative understanding of the subject. Therefore, the future research should concentrate on finding reasonable ways to quantify and measure the digital divide.

References

1. Noh, Y.-H., Yoo, K.: Internet, inequality and growth. J. Policy Model. **30**(6), 1005–1016 (2008)
2. Butcher, M.P.: At the foundations of information justice. Ethics Inf. Technol. **11**(1), 57–69 (2009)
3. Hilbert, M.: The end justifies the definition: the manifold outlooks on the digital divide and their practical usefulness for policy-making. Telecommun. Policy **35**(8), 715–736 (2011)
4. Van Dijk, J.A.: Digital divide research, achievements and shortcomings. Poetics **34**(4–5), 221–235 (2006)
5. Skok, W., Ryder, G.: An evaluation of conventional wisdom of the factors underlying the digital divide: a case study of the Isle of Man. Strateg. Change **13**(8), 423–428 (2004)
6. Billon, M., Lera-lopez, F., Marco, R.: Differences in digitalization levels: a multivariate analysis studying the global digital divide. Rev. World Econ. **146**(1), 39–73 (2010)
7. Tipton, F.B.: Bridging the digital divide in Southeast Asia: pilot agencies and policy implementation in Thailand, Malaysia, Vietnam, and the Philippines. ASEAN Econ. Bull. **19**(1), 83–99 (2002)
8. Olaniran, B.A., Agnello, M.F.: Globalization, educational hegemony, and higher education. Multi. Educ. Technol. J. **2**(2), 68–86 (2008)
9. Brooks, S., Donovan, P., Rumble, C.: Developing nations, the digital divide and research databases. Ser. Rev. **31**(4), 270–278 (2005)
10. Vodoz, L., Reinhard, M., Barbara, G.: Pfister, The farmer, the worker and the MP. GeoJournal **68**(1), 83 (2007)
11. Zhang, X.: Income disparity and digital divide: the internet consumption model and cross-country empirical research. Telecommun. Policy **37**(6–7), 515 (2013)

12. Quibria, M.G., et al.: Digital divide: determinants and policies with special reference to Asia. J. Asian Econ. **13**(6), 811 (2003)
13. Middleton, K.L., Chambers, V.: Approaching digital equity: is wifi the new leveler? Inf. Technol. People **23**(1), 4–22 (2010)
14. Cruz-Jesus, F., Oliveira, T., Bacao, F.: Digital divide across the European Union. Inf. Manage. **49**(6), 278–291 (2012)
15. Stanley, L.D.: Beyond access: psychosocial barriers to computer literacy. Inf. Soc. **19**(5), 407–416 (2003)
16. Cullen, R.: Addressing the digital divide. Online Inf. Rev. **25**(5), 311–320 (2001)
17. Antonelli, C.: The digital divide: understanding the economics of new information and communication technology in the global economy. Inf. Econ. Policy **15**(2), 173 (2003)
18. Qingxuan, M., Mingzhi, L.: New economy and ICT development in China. Inf. Econ. Policy **14**(2), 275 (2002)
19. Menou, M.J.: The global digital divide; beyond hICTeria. In: Aslib Proceedings. MCB UP Ltd (2001)
20. Mansell, R.: Constructing the knowledge base for knowledge-driven development. J. Knowl. Manage. **6**(4), 317–329 (2002)
21. Dewan, S., Frederick, J.R.: The digital divide: current and future research directions. J. Assoc. Inf. Syst. **6**(12), 298–336 (2005)
22. Drouard, J.: Costs or gross benefits? - What mainly drives cross-sectional variance in Internet adoption. Inf. Econ. Policy **23**(1), 127 (2011)
23. Edwards, Y.D.: Looking beyond the digital divide. Fed. Commun. Law J. **57**(3), 585–592 (2005)
24. Moss, J.: Power and the digital divide. Ethics Inf. Technol. **4**(2), 159 (2002)
25. Mordini, E., et al.: Senior citizens and the ethics of e-inclusion. Ethics Inf. Technol. **11**(3), 203–220 (2009)
26. Couldry, N.: New media for global citizens? The future of the digital divide debate. Brown J. World Aff. **14**(1), 249–261 (2007)
27. Tavani, H.T.: Ethical reflections on the digital divide. J. Inf. Commun. Ethics Soc. **1**(2), 99–108 (2003)
28. DiMaggio, P., et al.: Social implications of the internet. Annu. Rev. Sociol. **27**, 307–336 (2001)
29. Katz, J.E., Rice, R.E., Aspden, P.: The Internet, 1995-2000: access, civic involvement, and social interaction. Am. Behav. Sci. **45**(3), 405–419 (2001)
30. Fuchs, C., Horak, E.: Africa and the digital divide. Telematics Inform. **25**(2), 99–116 (2008)
31. Pick, J.B., Nishida, T.: Digital divides in the world and its regions: a spatial and multivariate analysis of technological utilization. Technol. Forecast. Soc. Chang. **91**, 1–17 (2015)
32. Cooke, L., Greenwood, H.: "Cleaners don't need computers": bridging the digital divide in the workplace. Aslib Pro. **60**(2), 143–157 (2008)
33. Pittman, J.: Converging instructional technology and critical intercultural pedagogy in teacher education. Multicultural Educ. Technol. J. **1**(4), 200–221 (2007)
34. Lee, J.-W.: The roles of demographics on the perceptions of electronic commerce adoption. Acad. Mark. Stud. J. **14**(1), 71–89 (2010)
35. Norris, P.: Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide. Cambridge University Press, Cambridge (2001)
36. Van Dijk, J., Hacker, K.: The digital divide as a complex and dynamic phenomenon. Inf. Soc. **19**(4), 315–326 (2003)
37. Kyriakidou, V., Michalakelis, C., Sphicopoulos, T.: Digital divide gap convergence in Europe. Technol. Soc. **33**(3–4), 265–270 (2011)

38. Vicente, M.R., López, A.J.: Assessing the regional digital divide across the European Union-27. Telecommun. Policy **35**(3), 220–237 (2011)
39. Rao, S.S.: Bridging digital divide: efforts in India. Telematics Inform. **22**(4), 361–375 (2005)
40. Schleife, K.: What really matters: regional versus individual determinants of the digital divide in Germany. Res. Policy **39**(1), 173 (2010)
41. Douglas, H.: The rural-urban digital divide. J. Mass Commun. Q. **77**(3), 549–560 (2000)
42. Novo-Corti, I., Varela-Candamio, L., García-Álvarez, M.T.: Breaking the walls of social exclusion of women rural by means of ICTs: The case of 'digital divides' in Galician. Comput. Hum. Behav. **30**, 497–507 (2014)
43. Bach, M.P., Zoroja, J., Vukšić, V.B.: Determinants of firms' digital divide: a review of recent research. Procedia Technol. **9**, 120–128 (2013)
44. Graham, S.: Bridging Urban digital divides? Urban polarisation and information and communications technologies (ICTs). Urban Stud. (Routledge) **39**(1), 33–56 (2002)
45. Peng, G.: Critical Mass, diffusion channels, and digital divide. J. Comput. Inf. Syst. **50**(3), 63–71 (2010)
46. Burkhard, R.A.: Learning from architects: complementary concept mapping approaches. Inf. Visual. **5**(3), 225 (2006)

Sentiment Analysis of Products' Reviews Containing English and Hindi Texts

Jyoti Prakash Singh[1(✉)], Nripendra P. Rana[2], and Wassan Alkhowaiter[3]

[1] National Institute of Technology Patna, Patna, Bihar, India
jyotip.singh@gmail.com
[2] School of Management, Swansea University, Swansea SA2 8PP, UK
n.p.rana@swansea.ac.uk
[3] Qassim University, Buraidah, Saudi Arabia
w.alkhowaiter@gmail.com

Abstract. The online shopping is increasing rapidly because of its convenience to buy from home and comparing products from their reviews written by other purchasers. When people buy a product, they express their emotions about that product in the form of review. In Indian context, it is found that the reviews contain Hindi text along with English. It is also found that most of the Hindi text contains opinionated words like *bahut achha*, *bakbas*, *pesa wasool* etc. We have tried to find out different Hindi texts appearing in product reviews written on Indian E-commerce portals. We have also developed a system which takes all those reviews containing Hindi as well as English texts and find out the sentiment expressed in that review for each attribute of the product as well as a final review of the product.

Keywords: Sentiment analysis · POS-Tagging · Review analysis · Product summarization

1 Introduction

The life style of society is changing with the penetration of Internet, and E-commerce in every corner of the world. Earlier, the advertisement and friends recommendations were a major source of information while buying a product. The number of recommendations was a limited one to compare similar products of different brands. Nowadays, as the e-commerce business has grown up, they are offering more products. The e-commerce websites also request their customers to write their experience about the product they brought in the form of a product review. These reviews offer significant information to buyers about the product they are planning to buy and also enable them to compare products of different brands. The reviews help consumers to choose the best products by comparing them based on other consumers' evaluation of the products. It also aids in the improvement of the product by informing the manufacturers about the advantages and defects of their products. The number of reviews about the

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 416–422, 2015.
DOI: 10.1007/978-3-319-25013-7_33

products grows with the growth of e-commerce businesses. It becomes very difficult for buyers and sellers to manually analyze a large number of reviews and get any meaningful information. This attracts a lot of researchers to automate the analysis of reviews and get valuable information hidden in the reviews [3,5].

The reviews written by Indian buyers are mainly in English, but it contains some Hindi texts (written in English Scripts only) also as Hindi is a prevalent language in India. Some of the most widely used Hindi words like *bahut achha*, *bakbas*, *pesa wasool* are found in a number of reviews. Most of these words are opinionated and contain strong opinions in the form of *good* or *bad*. Most of the earlier work done in the area of finding polarity of opinions for product reviews neglect these texts as they are mainly developed for English texts only. As per the best of our knowledge, no work has yet been reported which consider the correction of these typos and includes the sentiment of the Hindi words along with English texts.

In this work, we have proposed a sentiment analysis system which works for reviews containing both English as well as Hindi opinionated texts. First of all we have gathered possible Hindi opinionated texts from reviews appearing on Indian popular E-commerce sites such as *amazon.in, flipkart.com, snapdeal.com, shopclues.com* and so on. These Hindi texts are preprocessed and their equivalent English words are found. The summarized review of the product is then calculated consulting sentiwordnet database.

The rest of the paper is organized as follows: The proposed system architecture and algorithm are discussed in Sect. 2. In Sect. 3, we present our results and finally in Sect. 4, we conclude the paper.

2 Proposed Work

Product reviews contents from popular Indian e-commerce sites like *amazon.in, flipkart.com, snapdeal.com, shopclues.com* are collected as our dataset. The dataset has a lot of typos in the form of joint words like *verygood* as well as abbreviations containing numerals such as *gr8* for *great*. The dataset also contains Hindi words like *"bakbas", "bekar", "achchha"* (written in English script only). Some sample typos gathered from various Indian e-commerce websites are given in Tables 2 and 3. Table 1 contains words of Hindi Texts typed in English along with their English equivalent text. Table 2 contains some popular abbreviations used online for review, chatting, etc. along with their correct form in English. Some joint words (missing space) are shown in Table 3.

One of the primary focuses of this work is to pre-process the product review available on Indian E-commerce sites so that the reviews contain only English text. Once reviews are converted to English text, Part of Speech (POS) Tagging to the text is done using wordnet [4] database. Once POS tagging is done, the adjective, noun, and adverb are extracted. Further, sentiwordnet [1,2] database is used to assign numerical values to the adjectives contained in the review. The proposed system architecture is shown in Fig. 1.

Table 1. List of wrong words in Hindi

Wrong words	Corrected words
Ye	This
Achha	Good
G8t	Great
N8t	Night
H	Is
Som1	Some one

Table 2. List of wrong words in English

Wrong words	Corrected words
Gud	Good
Gooood	Good
Exclent	Excellent
Bd	Bad
Awesm	Awesome

A pseudo code for our proposed system is given below.

Proposed Algorithm:

Step 1: Tokenize based on space.
Step 2: Consult wordnet
If word is matched, then go to POS tagger.
else correct word and go to POS Tagger
Step 4: Noun, Adverb, and Adjective are stored in frequent feature database.
Step 5: Generate the product summary with the help of SentiWordNet Lexical databases.

The working of our scheme is traced with the aid of an example presented here. *Yeh achha camera h. eski pictre quality bahut achhi hai. The pics resolution is enough. Zoom is bakbas. focus is verygd bt not g8t.*

The document (Complete review) is broken down into several sentences based on [.], [?], And [!] Mark. For example review, sentences are:
S1. Yeh achha camera h.
S2. eski pictre quality bahut achhi hai.
S3. The pics resolution is enough.
S4. Zoom is bakbas.
S5. focus is verygd bt not g8t.

Yeh is a Hindi word whose English equivalent is *This*. *Achha* is another Hindi word meaning *good* in English. The complete review is written in English after correcting and converting every word to English.

Table 3. List of joint words

Jointed words	Corrected words
Verygood	Very good
Verybad	Very bad
Bahutbura	Bahut bura

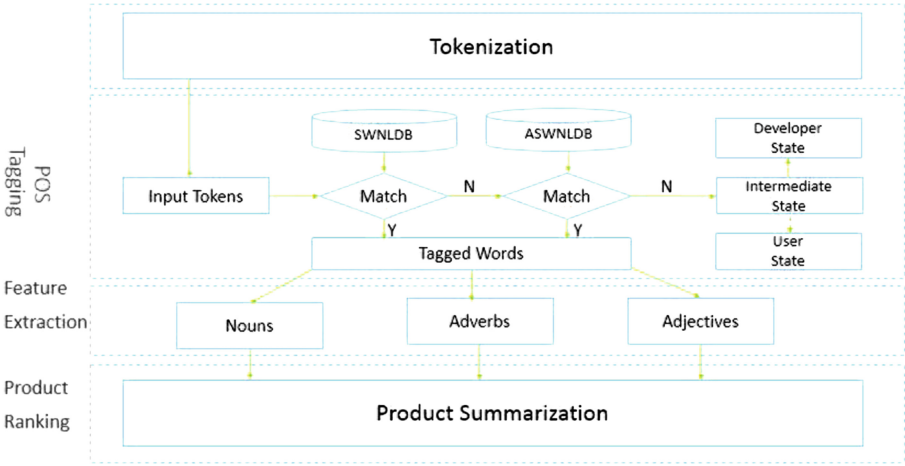


Fig. 1. Proposed system architecture

s1. This is a good camera.
s2. Its picture quality is very good.
s3. The picture resolution is enough.
s4. Zoom is bad.
s5. Focus is very good but not great.

The POS tagging is applied as given below for just one sentence. We have used Penn Treebank tagset for Part of Speech Tagging

This========== [This_DT]
Good========== [Good_JJ]
Camera========= [Camera_NN]
Is============ [Is_VBZ]
. ============[._UH]

Next step is to consult the sentiwordnet database to find the priority of the adjective to find the sentiment value of the sentence. The score of every adverb and adjective are given in the SentiWordNet lexical database. We have listed here some of them given in Table 4 which are going to be used in the above example.

Table 4. Score list of words

Word	Orientation of word	Score
Good	Positive	.75
Great	Positive	.875
Awesome	Positive	.875
Excellent	Positive	1
Well	Positive	.75
Average	Positive	.375
Enough	Neutral	.875
Bad	Negative	.65
Very	Nil	.5
Not	Negation	−1

Table 5. Sentence wise score

Sentence number	Score	Score type
S1	0.750	Positive
S2	1.250	Positive
S3	0.125	Positive
S4	0.625	Negative
S5	0.275	Positive

Where nil represents the neither positive nor negative orientation of words. And negation represents the multiplier factor which having −1 value.

For above example, sentence s1 has good as adjective whose sentiment score is +0.75. In this sentence there is no adverb or negation, so the sentiment score for sentence s1 is +0.75. For second sentence s2, adverb *(very)* is there with the adjective *(good)*, so the sentiment score of s2 is 1.25, which is a sum of scores of good (0.75) and very (0.5). The sentiment score of each sentence is shown in Table 5.

The polarity of the review of a product is determined by finding the polarity of each feature of the product across all reviews and finding a weighted sum of all features.

3 Result

We have collected 1100 reviews from *flipkart.com* and *amazon.in* of three popular mobile brands in India at the time of writing this paper. The results show that for Android based smart-phone people are talking about features like *camera, battery, memory, processor, RAM, display, price, weight and phone*. Out of these features, battery, camera and display are found to be more prominent ones across

Table 6. Summary of the review of product **Samsung Galaxy S3 Neo**, **Asus Zenfone 2** and **Honor 4X**

Phone feature		Galaxy S3 Neo	Zenfone 2	Honor 4X
Feature Name	*Scores Type*	*Percentage(%)*	*Percentage(%)*	*Percentage(%)*
Camera	Positive	100	66.39	60.17
	Negative	0.0	33.64	39.82
Battery	Positive	61.76	90.82	66.09
	Negative	38.23	9.18	33.9
Memory	Positive	100	100	49.24
	Negative	0.0	0.0	50.76
Processor	Positive	100	100	49.25
	Negative	0.0	0.0	50.74
RAM	Positive	100	100	48.24
	Negative	0.0	0.0	51.75
Display	Positive	90.5	0.0	100
	Negative	9.5	100	0.0
Price	Positive	0.0	No opinion	No opinion
	Negative	100	No opinion	No opinion
Weight	Positive	100	0.0	No opinion
	Negative	0.0	100	No opinion
Phone	Positive	55.07	89.09	62.02
	Negative	44.93	10.91	37.98
Overall	**Positive**	**68.86**	**75.14**	**62.60**
	Negative	**31.134**	**21.15**	**32.26**

all phones. The results are shown in Table 6. *no opinion* in both positive and negative rows shows that no one has given any opinion about that feature of that product.

4 Conclusion

We have designed a sentiment analysis system which can take reviews written in Hindi as well as English texts and find the sentiment of customers for that product. We have taken a dictionary based approach to correct the wrong words and replace Hindi text with their English equivalent. We further want to extend this system with a machine learning algorithm to correct the wrong words and Hindi words. The final opinion score is a weighted average of all the features of the product under consideration. We are also working to identify the most prominent features of a product to calculate the final opinion score.

References

1. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: LREC, vol. 10, pp. 2200–2204 (2010)
2. Esuli, A., Sebastiani, F.: Sentiwordnet: a publicly available lexical resource for opinion mining. In: Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006), pp. 417–422 (2006)
3. Hu, M., Liu, B.: Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD 2004, pp. 168–177. ACM, New York (2004)
4. Miller, George A.: Wordnet: a lexical database for english. Commun. ACM **38**(11), 39–41 (1995)
5. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. HLT 2005, pp. 339–346. ACL, Stroudsburg (2005)

Adaptive Normative Modelling: A Case Study in the Public-Transport Domain

Rob Christiaanse[1,2], Paul Griffioen[1], and Joris Hulstijn[1(✉)]

[1] Delft University of Technology, Delft, Netherlands
{p.r.griffioen,j.hulstijn}@tudelft.nl
[2] EFCO Solutions, Amsterdam, Netherlands
r.christiaanse@efco-solutions.nl

Abstract. Data analytics promises to detect behavioral patterns, which may be used to improve decision making. However, decisions need to motivated, and they are often motivated by models. In this paper we explore the interplay between data analytics and process modeling, specifically in normative settings. We look specifically at value nets, mathematical models of the flow of money and goods, as used in accounting. Such models can be used to analyze the proportions of various flows, such as resources consumed and products produced. Such analyses can be used in the planning and control cycle, for forecasting, setting a budget, testing and possibly adjusting the budget. In other words, it can be used for adaptive normative modeling. We look in particular at a case study of a provider of public transport services for school children. The case shows that the use of value nets for analysis of proportions is (i) feasible, and (ii) useful, in the sense that it provides valuable new insights about the revenue model.

Keywords: Data analytics · Business process modeling · Normative systems

1 Introduction

Data analytics is heralded as the new big thing [5]. People, services, goods or processes leave traces, producing vast amounts of data, which can be analyzed. Some have claimed about 'Big Data' that the sheer volume of data and the power of the analytical tools would remove the need for scientific method [2]. Patterns in behavior emerge which can be made actionable, i.e., they can used to make better decisions. We disagree. Decisions that matter must be motivated. In many application domains, professional decisions are based on models. In particular, this is true in normative settings, where situations are compared to norms or standards, which can be either based on practices and conventions, or on formal rules and regulations. In such normative settings, the model has become a reference model or norm. For instance, in compliance checking, a 'de jure' model, based on the law, is compared with a 'de facto' model, based on data about the real situation [10].

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 423–434, 2015.
DOI: 10.1007/978-3-319-25013-7_34

In this paper we explore the interplay between computational models, on the one hand, and data analytics tools on the other hand for decision making in a very specific task and domain. In particular, we focus on the planning and control cycle, that involves forecasting, setting a target, and comparing results to the target [12]. Clearly, this setting is normative and the decisions are guarded by a set of expectations and norms. In this setting there is also a challenge: data analytics is essentially empirical; it is exploratory and starts from the available data. Analyzing models in a computational setting, on the other hand, is typically based on theory, generalizations and abstractions. The challenge is to combine the power of data analytics, with the justification of models for decision making. Moreover, as the world changes, it is likely that models need to be adjusted. How can models be adapted? In other words: we need *adaptive normative modeling*. In Sect. 2 we argue that value nets model the essential process flow ratios that underly the numerical relation between data analysis and process modeling.

Value nets can be viewed as instantiations of business process models useful for process diagnosis and norm analysis. Norm analysis is a foundational notion buttressing the planning and control cycles within organizations. Just as data analytics, business process modeling is interested in finding and describing structural patterns in data. In this sense structural means that the decision structure is explicitly captured. Value nets provides in a computational framework for the modeling and analysis of business processes that focuses on the normative ratios in enterprise behavior for the purpose of quality management, quality assurance, and measurement. Here do data-analytics and value nets modeling meet.

In Sect. 3 value nets will be modeled as extended Petri Nets that capture the flow of values in an enterprise. From this model an enterprise's normative cyclic behavior can be computed as so called tours. These tours contain the ratios of the process events can be factored into what Ijiri calls causal chains of events. Section 4 shows how the causal chains can be computed by factoring the value net's incidence matrices. In the case study in Sect. 5 it is shown how this is used to interpret the data for planning and control.

The case study focuses on the planning an control cycle of a public transport provider, providing schoolbus services to schoolchildren. The governance setting is similar to that described in [6]. This is a highly regulated domain; compliance plays an important role. As you can imagine, this involves a lot of data about transport movements, which may contain many hidden pattens of behavior. Some of these patterns are relevant to the auditing task: are the invoices legitimate? Other patterns are more relevant to management, when they want to improve efficiency. We have access to this data.

The research can be positioned as part of Business Process Management (BPM) [4,7]. Here, we observe the same contrast in attitude between modeling (forward looking; engineering) and data analytics (backward looking; empirical). Related to this, we made three observations about the way BPM has developed, which make it less suitable for the management accounting domain. First, much of the research effort in BPM has traditionally been focused on process models: specification and verification of formal properties, such as termination or

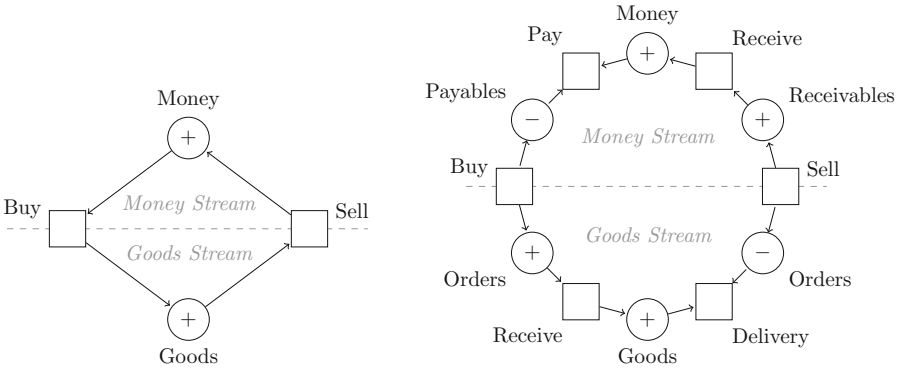


Fig. 1. Value Nets for trading companies. The horizontal dashed line splits the model into the *money flow* in the upper part and the *goods flow* in the lower part.

conformance to rules. This is done at design time. What has sometimes been lacking, is a look at real data, after execution, to see whether these generalizations apply in practice. Second, BPM models, in particular those based on work-flow, typically look at individual cases. However, in accounting and management, we are more interested in aggregated data: how many cases of a certain type per month? What was the use of resources? This aspect is also taken on by process mining. Third, BPM models tend to focus on the existence of dependencies between process steps or activities; we are just as much interested in the amount of value or the number of objects flowing along these links. In other words, we are not only interested in the control flow, but rather in the size of the flows.

The contribution of this paper is a computational framework for the modeling and analysis of business processes that focuses on the normative ratios in enterprise behavior for the purpose of quality management, quality assurance, and measurement. These normative ratios provide in the de jure models which can be compared with the de facto models extracted from the data using data analytics and vise versa.

2 Business Process Modeling and Data Analytics

The price mechanism ensures that a market price of a good and or service accurately summarizes the vast array of information held by market participants [3]. It needs no elaboration that the information aggregation characteristic of the pricing system buttresses many theories about the communicative function of prices and the decisions participants in the marketplace make to exploit business opportunities in the creation of value.

Management of enterprises decide upon a business model that depicts the transaction content, structure, and the governance designed so as to create value through the exploitation of business opportunities" [1]. The transaction content

refer to goods, services or the information exchanged, where the transaction structure defines the way parties i.e. agents participate in the exchange and how they are (inter)linked. Transaction governance refers to the legal form of organization, and to the incentives for the participants in transactions. Hence data, processes and governance are intertwined notions which are strongly related to each other.

More specifically a contract is an instantiation of a business model. Consequently a contract depicts the agreed upon content, structure, the incentives and the rules of conduct among parties involved in the contract. In the case a contract is executed the buyer receives goods or services from the vendor in the agreed upon quality and the buyer pays the vendor for the agreed upon price coined as value. Actually aforementioned paying mechanism is often referred to as the revenue model of the vendor. More specifically a revenue model refers to the specific modes in which a business model enables revenue generation.

Economic transactions as described above can be formally modeled as a value cycle which is interlinked to the business model of the enterprise i.e. the revenue model. This view is inspired by accounting models in the owner ordered accounting tradition [16] and value chain theory [13]. When used in computer science, often the purpose of these models is to analyze the representations of actions and events in a business process, and study their well-formedness. Consequently an intra-organizational workflow i.e. business processes can be modeled as a value net which provides a top-level view of an enterprise that focuses on the economic events equivalent to the value cycle of an enterprise. Figure 1 shows an elementary and a more elaborate example of a value cycle for a trading company. Each event is a transfer of value.

3 Value Cycle and Value Nets

A value net is modeled with a dimensioned Petri Net, extended with a place sign and a valuation, with some special structural characteristics that make it a good representation of the intra-organizational structure. Figure 2 shows the value cycle for the case study from the next sections.

Definition 1. *A* Value Net *is a tuple* (P, T, F, B, s, v) *with*

- P *a finite collection of places,*
- T *a finite collection of transitions,*
- $F, B : P \times T \to \mathbb{N}$ *the net's incidence matrices*
- $s : P \to \{asset, liability\}$ *the indicator for each place's sign, and*
- $v : P \to \mathbb{R}$ *a valuation.*

Additionally a value net has the following structural properties:

- The net is cyclic in the following sense. There is a special place labeled *money* from which it is possible to reach every other node in the model and that can be reached from every node in the model. Put differently, every node in the model is on a path from *money* to *money*.

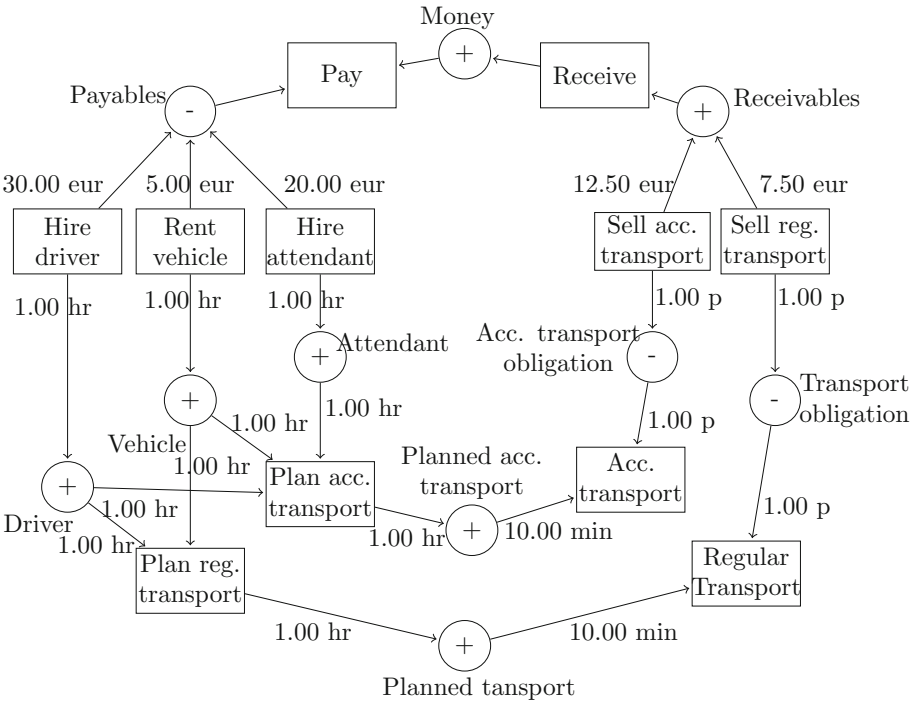


Fig. 2. Value Net for a transport provider.

– The net is divided in a money part and a production part. Place *money* is in the money part. The transitions in which the input places are in the money part and the output places in the production part are called *buy* transitions. The transitions in which the input places are in the money part and the output places in the production part are called *sell* transitions.

We use subscripts to differentiate the money and the production part. So, partitioning $P = P_{mon} \cup P_{prod}$ splits the places into the money part and the production part. Transitions in T_{mon} are between places the money part and transitions in T_{prod} between places in the production part. A T_{buy} transition consumes value from the money part and produces value in the production part. A T_{sell} transition consumes value from the production part and produces value in the money part.

A value net's semantics follows from the Petri Net semantics. A marking m is a $P \rightarrow \mathbb{N}$ vector that denotes the number of tokens in each place. A transition vector t is enabled when $F \cdot t \leq m$. When it fires it consumes $F \cdot t$ tokens and it produces $B \cdot t$ tokens. Let flow matrix G be defined by $G = B - F$. The effect of transition vector t is then the marker change computed as the matrix-vector product $G \cdot t$.

A value net extends a Petri net with a valuation, because for the computation of (normative) behavior in workflows it is essential to distinguish in monetary

units and in product units. Just working with values leads to information loss, because value is the product of a valuation and a quantity.

$$\text{Value} = \text{Valuation} \times \text{Quantity}$$

A value net in monetary units has the advantage that everything is commensurable. It is even a requirement for the balancing of events in double bookkeeping [8,11,14]. However, it is unsuitable for conservation laws because the value depends on a valuation that fluctuates. As soon as money is exchanged for a good this good gets exposed to price fluctuations. Only when the good is exchanged into money again later the effect on value becomes fixed. Exact connections in the product part is only feasible in product units. Therefore it is necessary to split values into valuations and quantities and model a business model in product units. A valuation is a linear property of a marking with a monetary unit of measurement. It gives the total economic value of the tokens in a dimensioned net. A weight vector v contains the value of a token for each place. Given a marking m the inner product $\langle v, m \rangle$ is the total value.

The other extension is the distinction between assets and liabilities. Each place in a value net represents either a positive or a negative value. In a value net the positive values flow anti-clockwise and the negative values flow clockwise. In many computations it is convenient to have all value flow in the same direction, and in these case we use an unsigned value net. Let matrix G be the flow matrix of a value net and let matrix S be a diagonal matrix with $+1$ for positive and -1 for negative places, then matrix $U = S \cdot G$ is the *unsigned flow matrix*, and the corresponding value net is called the *unsigned value net*. In an unsigned net the edges connected to a negative place are switched with respect to the signed net. An unsigned value net produces the same outcome but with signs switched. Multiplying with sign matrix S again gives the identical outcome.

4 Causal Chains of Events and Reconciliation Relations

A tour relates the behavior of an enterprise to the cyclic structure of its value net i.e. a process model as depicted in Fig. 2. The cyclic structure is not directly apparent in behavior because transitions can fire at any point in the cycle. We want to group transactions that make up what Ijiri calls a causal chain of events [11]. In general terms such a chain of events is a variation on buying products and resources, producing an end product, and selling the product. The cycle starts with the consumption of money and ends with the production of money. Such cyclic behavior follows from the cyclic structure of a business process.

A formal notion of the tour concept was introduced for an audit context in [9]. It defines a tour as a constellation of events whose total effect is on money only. All other produced tokens are at some point consumed by another step in the process.

Definition 2. *A transition vector t with firing result $m = G \cdot t$ is a tour when $m[\text{money}] > 0$ and $m[x] = 0$ for any place $x \neq \text{money}$.*

Place	$F_{\text{buy}} \cdot t$	$B_{\text{buy}} \cdot t$	$C \cdot t$	$F_{\text{sell}} \cdot t$	$B_{\text{sell}} \cdot t$
payables	3500 c	-	-	-	-
vehicle	-	3600 s	-	-	-
driver	-	3600 s	-	-	-
planned reg. transport	-	-	3600 s	-	-
reg. transport obligation	-	-		6 p	-
receivables	-	-	-	-	4500 c

Place	$F_{\text{buy}} \cdot t$	$B_{\text{buy}} \cdot t$	$C \cdot t$	$F_{\text{sell}} \cdot t$	$B_{\text{sell}} \cdot t$
payables	5500 c	-	-	-	-
vehicle	-	3600 s	-	-	-
driver	-	3600 s	-	-	-
attendant	-	3600 s	-	-	-
planned acc. transport	-	-	3600 s	-	-
acc. transport obligation	-	-		6 p	-
receivables	-	-	-	-	7500 c

Fig. 3. Factored tour for the transport case from Fig. 2. From left to right the table shows the sacrificed value, the purchased products, the intermediate products, the produced products and the received value.

Vector t is the number of times an events has to occur per tour. Every time such a tour has occurred the amount of money increases by m[money]. This increase is called the value jump. This formal tour concept captures the essence of the concept of value form a enterprise point of view.

With the unsigned value net we can reveal the various parts of a tour by factoring tour result $S \cdot U \cdot t$.

$$U = U_{\text{mon}} + U_{\text{buy}} + U_{\text{prod}} + U_{\text{sell}} \tag{1}$$

If we also split matrix U into $B - F$ to see the difference between token production and token consumption then we have the factoring to express connections between the value cycle's parts. Let t be a tour. Unsigned tour result $U \cdot t$ is factored in a causal chain as follows

$$(B_{\text{mon}} + B_{\text{buy}} + B_{\text{prod}} + B_{\text{sell}} - F_{\text{mon}} - F_{\text{buy}} - F_{\text{prod}} - F_{\text{sell}}) \cdot t \tag{2}$$

The various terms in this expression are the input and output of the different steps in the causal chain.

Grouping the money part and the production part immediately give the following equations:

$$(B_{\text{buy}} + B_{\text{prod}} - F_{\text{prod}} - F_{\text{sell}}) \cdot t = 0 \tag{3}$$

$$(B_{\text{sell}} + B_{\text{mon}} - F_{\text{mon}} - F_{\text{buy}}) \cdot t = G \cdot t \tag{4}$$

The matrix in $(B_{\text{sell}} + B_{\text{mon}} - F_{\text{mon}} - F_{\text{buy}}) \cdot t$ is exactly the flow matrix filtered for the production places and does not contain the money place. Because t is a

tour, the product must therefore be zero according to Definition 2. The second equation then follows immediately because the tour result $G \cdot t$ must result from the other parts of Eq. 2.

The first equation is Kirchhof's law for the production process. It shows that in a tour the input and output of production cancel. The second equations shows the value jump. It is the difference between monetary value produced and consumed.

Figure 3 shows the chains of events for the transport service provider from the case study. In this case there are two tours; one for regular transport, and one for accompanied transport. The case study focuses on the tour for regular transport.

5 Case Study: Public Transport Services

In 2010 a transport provider has agreed to provide so called taxi-bus transport services for secondary school pupils, who need to attend special schools, which are at a distance. For transport of this group of passengers, many rules and regulations apply. In the request for proposals, the municipalities have laid down a number of requirements for the contract, concerning maximal waiting times, durations, and routing. Every month, the transport provider sends an invoice with a data file, detailing the number of trips, passenger details, routes, departure and arrival times, departure and arrival addresses, ordered and canceled trips etc. This data file provides evidence of the services delivered for the invoice to be paid by the municipalities.

The contract specifies the following revenue model. Hence a revenue model refers to the specific modes in which a business model enables revenue generation. Based on trip requests received from the municipality specifying travel specifications from A to B at time t, at week day w, the routing software package used by the transport provider calculates the daily routes at each time of day, entailing trip requests i.e. the number of school pupils per route. Parties agreed that the routing software package is used to predict the best route and duration, for the sake of the contract. This prediction is based on standard speeds (100 km/h for motorways, 70 km/h for main roads, 40 km/h in town etc.). For each route, this gives occupied vehicle hours, which is priced at a certain tariff: revenue per route = price per occupied hour * occupied vehicle hours. This completes the description of the revenue model.

The price per occupied hour is the result of a normative decision made by management of the transport provider offering their quote in competing for the contract in 2010. No need to say that the quotation is a best estimate of expected outcome for providing future transport services. Within the context of the planning and control cycle of the transport provider the actuals i.e. the realized data are gathered on a daily basis for monitoring purposes. Deviations to the norm are very interesting from an operational perspective because operational management learns about the quality of planning operations and more over whether the norms used in calculating the price per occupied hour is correct. From a

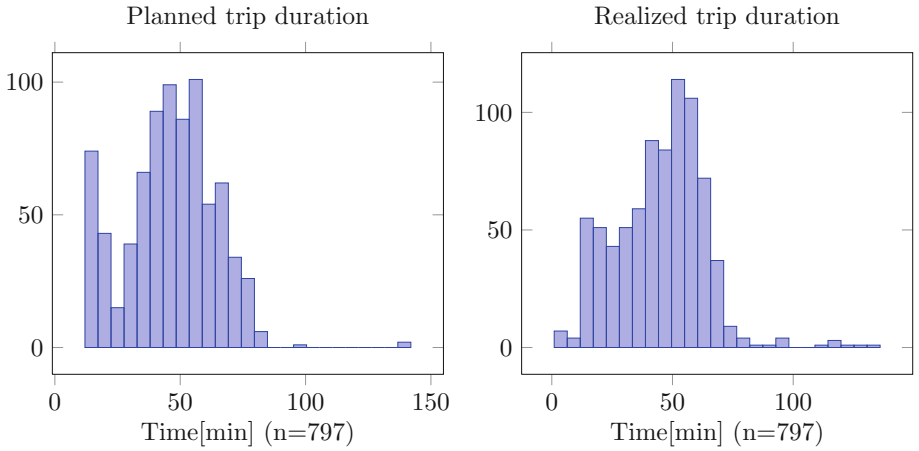


Fig. 4. Planned and realized trip durations. The average planned duration is 46.42 min, and the average realized duration is 45.10 min.

financial perspective the same information is very valuable in assessing whether the contract is profitable or not and learns the quality of the normative decision processes made by management.

In Fig. 2 the original value net buttressing the initial price bid is depicted. Factorization of the tour of the transport provider gives us the causal chain of events representing the normative representation of the inflows and outflows of business processes of the transport provider, revealing their normative ratios i.e. KPIs as input measures for analyzing deviations to the normative ratios. Hence on the right hand side we see that the sell of regular transport i.e. one trip from A to B at time t, at week day w, leads to a transport obligation. This transport obligation leads to planned transport measured in one hour. The value net shows that one trip equals 10 min drive so the occupancy rate is easily derived from these ratios. On average six trips make up one planned vehicle route. In general the occupancy rate of a vehicle is a key performance indicator for analyzing operational planning effectiveness, analyzing profitability of the contract at hand, and for most whether norms actually used in calculating pricing bids are sound. We refer to Fig. 3 for detailed exposition of the factored tours. Note that the value net is the instantiation of the revenue model of the contract and reveals the value creation within organizations.

We have access to the data detailing the number of trips, passenger details, routes, departure and arrival times, departure and arrival addresses, ordered and canceled trips etc. The period we studied comprised one month period. In Table 1 we give some preliminary figures of the data file en the computational results applying the value net logic as explained in section three and four on the planned data and realized data in that month. The occupancy rate is the reciprocal of the KIP, converted to hours.

The occupied vehicle hour planned and realized are depicted in Fig. 4. The average per route differs only slightly from each other, here 1.32 min. When we

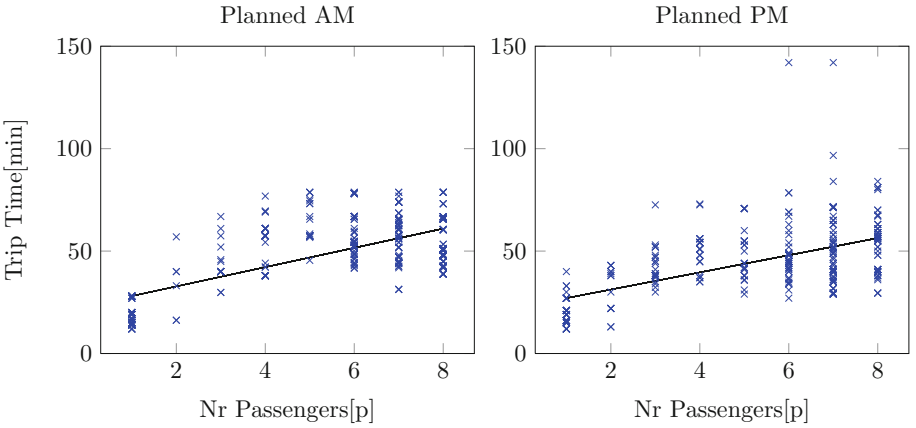


Fig. 5. Scatter plots for the planned trip durations in the morning (AM) and in the afternoon (PM).

Table 1. Comparison between planned and realized route durations, in the morning and in the afternoon.

	Planned AM	Planned PM	Realized AM	Realized PM
Mean	48.46 min	47.36 min	43.88 min	42.28 min
Std dev	17.57 min	19.06 min	18.00 min	16.83 min
KPI	9.07 min/p	8.86 min/p	8.73 min/p	8.41 min/p
Occupancy	6.62 p/hr	6.77 p/hr	6.87 p/hr	7.13 p/hr

combine this insight with the KPI and the occupancy rate then we may conclude from a financial and operational perspective that we have met our objective(s). Nothing has to be done. But if we look more closely to the data and make a more detailed analysis by splitting up the data in routes performed in the morning and in the afternoon than we must revise our opinion. The morning routes are more problematic as we thought. The overall performance in the afternoon is certainly better compared with the morning routes. The value net analysis gives us a different tour results. In this case financial management may still be content with the results but operations has now a motivation to look into the data asking why the overall performance in the afternoons is better than the morning routes. If we take a closer look at the planned data AM depicted in Fig. 5 and compare them with the realized data AM depicted in Fig. 6 than we see that the scatter plot of the KPI minutes per person per route is quite informative. In the case we compare these data with the planned and realized data PM than there are two things that are noteworthy. First we see that the morning routes show similar patters as the afternoon. After analyzing these data it turned out that traffic jams caused delays in the morning. Secondly after analyzing daily patterns it turned out that only two days in the week caused real delays. The planning

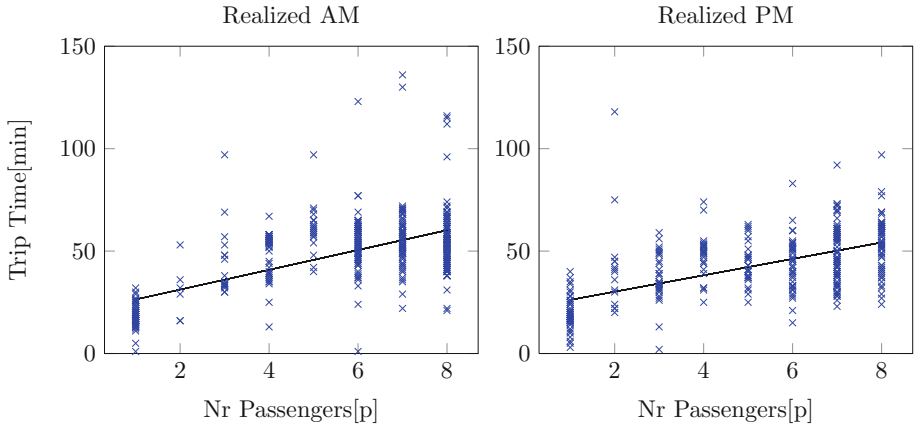


Fig. 6. Scatter plots for the realized trip durations in the morning (AM) and in the afternoon (PM).

department did not optimize routes for each route per day part. Rationally we would expect that the schedules were altered. In practice the schedules were kept on a weekly (average) basis. The value net analysis revealed that when we analyze the tours that actually one trip contains the following information: a school pupil travels from A to B at time t, at week day w. So the unit of analysis is a weekday, at time t, per school pupil which gave the correct information.

6 Conclusion

In the introduction we mentioned three aspects of BPM, which make it less suitable for the financial domain: (i) little empirical verification of applicability of models, (ii) focus on individual cases, instead of aggregate flows, (iii) focus on causal links, neglecting how much value or how many goods flow over the links.

There is an analogy between our approach and process mining [15]. Process mining also combines the exploratory and empirical aspects of data mining, with the use of linear models of processes, to analyze their formal properties. Process mining can – to a certain extend – also deal with the size of flows and use of resources. However, what it can't do is express the amount of value or number of object that flow or should flow over a link. That means that in normative settings, also process mining is left to analyze whether a process follows the procedure; it can't identify deviations in the (financial) content of transactions.

In Sects. 3 and 4 we elaborated on a computational approach for analysis of business processes models that focuses on the normative ratios in enterprise behavior for the purpose of quality management, quality assurance, and measurement. As shown in the case analysis the value net approach provides in the de jure models which can be compared with the de facto models extracted from

the data using data analytics and vice versa. This approach seems very fruitful for modeling adaptive mechanisms buttressing planning and control cycles within organizations.

References

1. Amit, R., Zott, C.: Value creation in E-Business. Strat. Manag. J. **22**(6–7), 493–520 (2001)
2. Anderson, Chris: The end of theory: the data deluge makes the scientific method obsolete. Wired Mag. 16(7) (2008)
3. Atakan, A.E., Ekmekci, M.: Auctions, actions, and the failure of information aggregation. Am. Econ. Rev. **104**(7), 2014–2048 (2014)
4. Cardoso, J., van der Aalst, W.M.P.: Handbook of Research on Business Process Modeling. Information Science Publishing, Hershey (2009)
5. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**(4), 165–1188 (2012)
6. Christiaanse, R., Hulstijn, J.: Control automation to reduce costs of control. Int. J. Inf. Syst. Model. Des. **4**(4), 27–47 (2013)
7. Dumas, M., van der Aalst, W., ter Hofstede Arthur, H.M.: Process-Aware Information Systems. Wiley, New Jersey (2005)
8. Ellerman, D.P.: Economics, Accounting, and Property Theory. Lexington Books, Lexington (1982)
9. Elsas, P.: Computational auditing. Ph.D. thesis, Vrije Universiteit Amsterdam (1996)
10. Governatori, G., Sadiq, S.: The journey to business process compliance, pp. 426–445. IGI Global (2009)
11. Ijiri, Y.: The Foundations of Accounting Measurement: A Mathematical, Economic, and Behavioral Inquiry. Prentice-Hall International Series in Management. Prentice-Hall, New Jersey (1967)
12. Merchant, K.A.: Modern Management Control Systems. Text and Cases. Prentice Hall, New Jersey (1998)
13. Porter, M.E.: CompetitIve Advantage: Creating and Sustaining Superior Performance. Free Press, New york (1985)
14. Rambaud, S.C., Pérez, J.G., Nehmer, R.A.: Algebraic Models for Accounting Systems. World Scientific, Singapore (2010)
15. Rozinat, A., van der Aalst, W.M.P.: Conformance checking of processes based on monitoring real behavior. Inf. Syst. **33**(1), 64–95 (2008)
16. Starreveld, R.W., de Mare, H.B., Joëls, E.J.: Bestuurlijke informatieverzorging. 1: Algemene grondslagen, vol. deel, 2nd edn. Samson Uitgeverij, Aplhen aan den Rijn/Brussel (1988) (in Dutch)

Business Process as a Service (BPaaS)

Model Based Business and IT Cloud Alignment as a Cloud Offering

Robert Woitsch(✉) and Wilfrid Utz

BOC Asset Management GmbH, Vienna, Austria
{robert.woitsch,wilfrid.utz}@boc-eu.com

Abstract. Cloud computing proved to offer flexible IT solutions. Although large enterprises may benefit from this technology by educating their IT departments, SMEs are dramatically falling behind in cloud usage and hence lose the ability to efficiently adapt their IT to their business needs. This paper introduces the project idea of the H2020 project CloudSocket, by elaborating the idea of Business Processes as a Service, where concept models and semantics are applied to align business processes with Cloud deployed workflows. Four architectural building blocks proposed for (i) design, (ii) allocation, (iii) execution and (iv) evaluation are discussed before providing and outlook.

Keywords: Business process as a service · Business processes in the cloud · Business and IT alignment · Meta modelling and semantic

1 Introduction

Cloud Computing is undoubtedly the current mega trend that has the potential to massively influence current use of IT, especially for business applications. Estimated improvements caused by efficient, flexible and networked IT resources range up to 30 % [1]. Hence, Cloud Computing is a chance for start-ups and smart companies that enter this global IT marketplace and obviously a risk for those who do not appropriately take advantage of Cloud Computing.

The ultimate challenge is to overcome so-called business and IT alignment, and bridge the gap between the business and IT domains. With respect to cloud offerings, this means that the current application view needs a corresponding business process view. Currently, typical parameters for SaaS – which is regarded as the current and upcoming market - are of a technical nature such as pricing models considering technical parameters, computing power, availability or network capacity. Business parameters such as legal aspects, business packages, process interoperability, or avoidance of vendor lock are used for distinction between different market players. Business Domain specific parameters like customer relationship for SMEs in the health domain, or web-appearance of an IT company are potential future options.

Hence we observe the need to abstract parameters from pure technical distinctions up to business and domain specific characteristics, which describe and distinguish cloud offerings.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 435–440, 2015.
DOI: 10.1007/978-3-319-25013-7_35

In the following, the approach of the EU project CloudSocket [2] for tackling the aforementioned issues is introduced.

2 BPaaS Use Cases

Our primary targets are SMEs that are currently excluded from using the cloud due to a lack of competence and high entry barriers. There is a gap between pragmatic, legally influenced and well-defined business processes, which are understood by SMEs, and a gigantic cloud market with numerous offerings that rarely consider the business episodes of an entrepreneur but focus on technical details. Startups and SMEs typically focus on their core business. Hence, there are several business processes such as customer relations and advertising, administrative issues on registration, IT services as well as after sales support that are necessary for business success, but can only be insufficiently supported by the IT resources of those organizations.

For a complete analysis of the use cases, please refer to [3].

2.1 Business Incubator Use Case

The Business Incubator focuses on supporting the "Coaching and Finance" efforts of start-ups facilitating designing, analyzing and simulating individual business plans and processes. These aspects also demand a high degree of adaptability of Cloud Services for Start-ups, e.g. Customer Relationship Management, Order Management, Human Resources Management.

Ecological Agriculture: A 28 year old biologist has an idea to take biological waste from a restaurant and stimulates a biological decomposition process. Usually such a process takes several years but the idea of the startup is to use worms to speed up this process.

- Initial situation: The startup presented the ideas to the business incubators. After this, the consultants have discussed with her about how to transform this business idea into a solid business model.
- CloudSocket technology intervention: The startup may require a range of different BPaaS, especially for customer relationship and worm production management.

Green Energy: This startup is a small-scale virtual power plant which connects to a grid infrastructure with power generation from wind, photovoltaic, and biogas. The company serves its customers with environmentally friendly energy for households and provides smart home functions through its remote access capability for turning appliances on or off.

- Initial situation: The company is intending to expand its services to include mobile energy sources for recharging electric cars and offer them for rental as a range-extension for drivers e.g. for a long weekend trip. The startup contacted the business incubator consultants.

- CloudSocket technology intervention: The startup may require a range of different BPaaS especially for customer relationship, partner management and internal management processes.

The observation in the first use case – the Business Incubator - is that supportive business processes can be applied across several startups. So BPaaS addressing e.g. Customer relationship can be offered to a wide range of startups.

2.2 Cluster Process Broker Use Case

The Business Process Broker use case identifies typical business episodes of potential SMEs in different application domains such as eHealth, Manufacturing, Photonics, Government, Security, e-Commerce, Retail, etc. but share a common set of business processes.

Internet Research and Procurement Process: An SME sells software and integrated appliances/electronic components that make devices "internet ready" in a few seconds.

- Initial situation: The SME continuously verifies prices of the electronic and mechanical components in the market and buys only products that match specific requirements in terms of customer needs and pricing. Monitoring the prices and the quality is a costly activity, which requires an ongoing analysis and trade-off between quality and price.
- CloudSocket technology intervention: The Company needs a solution that reduces the costs for procurement activities by improving the effectiveness. Generic self-management infrastructures or specially designed research processes including crawler and result databases have the potential to run in the cloud and to raise the productivity of this SME.

Kiosk Distribution Process: A company aims at distributing newspapers and magazines to kiosks and other points of sales in an Italian town. Every day, around 250 different Italian and foreign newspaper are delivered to 600 points of sales.

- Initial situation: Current customers are small kiosks with very limited IT infrastructure. Often the orders are realized via Facebook comments. To improve the maturity of the ordering and interaction process with those points of sales, a new but still lightweight Web-application must be provided.
- CloudSocket technology intervention: A new order process can be handled in the cloud, without IT installation on either the supplier or consumer sides. Furthermore the process can reflect a better understanding of the customer needs.

The observation in the second use case – the Cluster Broker - is that most of the potential end users of the CloudSocket have the potential need for a generic business process but also for more specific business processes. Hence the flexible configuration of business processes, hiding the complexity of the cloud and providing easy to use solutions, is a promising market potential.

2.3 The SME End User Perspective

In addition to the two aforementioned use cases – that describe the targeted end users market – we identify the steps of interested end users in a process-oriented approach.

We propose three steps for a typical SME as an end user.

- Check Cloud Readiness
- Transform Business Processes to be executable in the Cloud
- Enter the marketplace to access BPaaS

The project provides a checklist for SMEs and start-ups in order to check if they are capable of entering the cloud with their business processes. This framework is provided as an online questionnaire relating to business processes, see [4].

The transformation of business processes to be executable in the Cloud is divided in two transformations, whereas the first transformation is a horizontal one that transforms from one business process to another one and the second transformation is a vertical on transforming from business process to workflows. Although both business processes are not executable, the latter one has clear anchor points, where cloud offerings make sense. Hence the horizontal transformation extracts those parts of the process where a cloud offering can actually be applied.

The next transformation is a vertical one that maps to an executable workflow in the cloud - this actually provides the cloud offerings and enables the execution in the cloud. This next step is performed by entering the market place and selecting the most appropriate workflow that runs in the cloud. This selection can be supported by smart mechanisms.

2.4 The CloudSocket Broker Perspective

The so-called CloudSocket is a market place, where BPaaS are offered in a similar way to how SaaS are offered. Hence, from a market place point of view, the same mechanisms can be applied as for SaaS, as each "executable business process" can be identified as a workflow in the cloud with a corresponding end point. The difference between a SaaS and a BPaaS is the description in the form of a business process and its semantic annotation for enabling smart selection.

For organizations aiming for becoming a CloudSocket broker we propose: Plan, Build, Run, Check.

"Plan Business Processes" denotes the use of business process management tools to acquire, design, analyse and simulate and finally release domain-specific business processes. Here we understand business processes as a know-how platform of an organisation, hence those processes have the potential for domain-specific consultancy and improvement. Traditional business process management tools such as ADONIS® [5] are used.

"Build Business Processes" denotes that each of the aforementioned business processes are made executable by a set of deployable and executable workflows. We agreed to use the term workflow for processes that are orchestrated and executable on an IT platform to strengthen the difference to human orchestrated or executed business processes. Traditional workflow design tools like yourBPM [6] may be used.

"Run Business Process" indicates the provision and operation of processes as a service within a market place that are executed and run across services offered in the cloud. Although this is technically the most challenging part, the focus of this paper and the focus of the introduced CloudSocket project is on the alignment, hence the mapping between domain specific business processes and cloud deployable and executable workflows.

"Checking Business Processes" indicates the abstraction, using conceptual models and semantic, to introduce a semantic meaning into the purely technical data and process logs from the execution environment in the cloud. The meta model platform ADOxx [7] will be used to develop conceptual and semantic models that can be analysed and mapped to business processes.

3 BPaaS Reference Architecture

CloudSocket will provide business solutions to SMEs, which can be offered in an open and interoperable form. A particular focus is on startups which do not want to invest in their own IT infrastructure but concentrate on the development of their business. IT services need to be adapted according to the changes and evolution of the organization and business.

CloudSocket comprises four phases, each phase supported by a corresponding building block: (a) the design environment to describe business processes, business requirements and workflows (b) the allocation environment linking deployable workflows with concrete services, (c) the execution environment that executes and monitors the workflow as well as (d) the evaluation environment that lifts key performance indicators back to the business level.

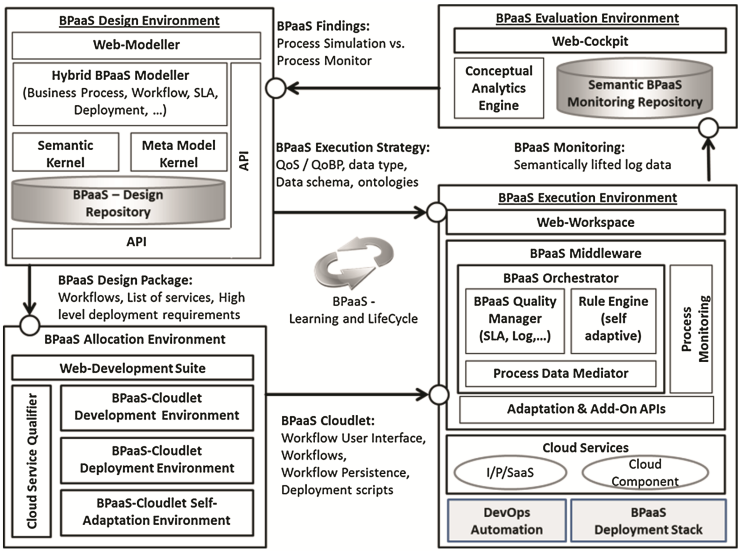


Fig. 1. CloudSocket high level architecture

Figure 1 introduces the four major building blocks, whereby each of the four building blocks supports one phase of the BPMS paradigm when applied for business process management in the cloud. The BPMS is a well-established business process management system paradigm that can be applied also on business process management in the cloud. More information on the architecture is available at [8].

4 Outlook

This paper introduces the idea of BPaaS in the context of the H2020 project CloudSocket, which started 01.01.2015. Hence this paper introduces the project idea of BPaaS and the use of a model-based approach to align domain specific business processes with cloud related executable workflows. First findings in identifying business episodes and possible business process models have been introduced and the current status of the architecture is briefly presented. By the time of the conference, the project can provide the first set of business episodes, a first reference architecture and an environment to check the cloud readiness of an end user.

All prototypes are available either as open source, open use or as provided services. A roadmap for technology provider indicates how alternative tools can be provided for a CloudSocket broker. Results on smart mapping domain specific business processes and cloud based executable workflows are expected by the end of 2015.

References

1. Liebhart, D.: Die modernen drei Musketiere, Cloud Computing, Computerwelt, März (2015)
2. CloudSocket Project. www.cloudsocket.eu. Accessed 18 May 2015
3. CloudSocket Project: D2.1 use case analysis. www.cloudsocket.eu. Accessed 18 May 2015
4. CloudSocket Project: D2.3 cloud transformation framework. https://cloudsocket.eu/transformation/. Accessed 24 July 2015
5. ADONIS Community Edition. www.adonis-community.com/. Accessed 24 July 2015
6. yourBPM. activiti.org/download. Accessed 24 July 2015
7. ADOxx. www.adoxx.org. Accessed 24 July 2015
8. CloudSocket Project: D4.1 First CloudSocket Architecture. www.cloudsocket.eu. Accessed 04 September 2015

Witness Workshop

IT-Enabled Resilient, Seamless and Secure Global Supply Chains: Introduction, Overview and Research Topics

Bram Klievink[1(✉)] and Gerwin Zomer[2]

[1] Delft University of Technology, Delft, The Netherlands
a.j.klievink@tudelft.nl
[2] TNO Netherlands Organisation for Applied Scientific Research, Delft, The Netherlands
gerwin.zomer@tno.nl

Abstract. This paper is the introduction to the fourth Workshop on *IT-enabled Resilient, Seamless and Secure Global Supply Chains* (WITNESS 2015). In the paper, we present a synthesis of the innovation agendas of a series of international research, development and demonstration projects that seek to make supply chains more efficient and more secure. From this agenda, we highlight three main innovation areas that are central to the current transition in the domain of logistics and international trade. Against this background, we introduce the five papers that are part of the WITNESS 2015 workshop. The papers primarily cover IT related innovations and include topics such as data quality and data governance, the development and interoperability of platforms that together present an IT innovation in international trade, and serious gaming for making key players aware of the potential of the innovations. We finalise with the suggestion to put more emphasis on the non-technical parts of the innovations, as that is what is ultimately needed to ensure wide uptake of the innovations, in order to make them really count.

Keywords: Information infrastructure · International trade · Collaboration · Innovation · Supply chains · Security · Information society · WITNESS

1 Introduction

Due to globalisation and automation, international supply chains have become more dynamic and more complex over the recent years. The commercial transactions, organisation of logistics and execution of transport are performed by multiple layers of organisations, all connected by contracts of sale and carriage [1]. Furthermore, as a consequence of efficiency developments like the introduction of sea container transports and consolidation of cargo, more and more goods are concealed from supply chain visibility, negatively impacting the ability of businesses and governments to assess the security of internationally shipped cargos [2]. Concepts like outsourcing of activities, consolidating cargo from multiple shippers and transporting across multiple modes of transport (sea, air, road, rail, river) have complicated the organisation and control of the chain and demand more from the involved companies in terms of collaboration and information sharing.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 443–453, 2015.
DOI: 10.1007/978-3-319-25013-7_36

In this context, enhancing supply chain security whilst preserving trade- and logistics-related economic growth, and at the same time safeguarding societal values, is a major topic on the agendas of global and national players, both public and private. To achieve this, various projects, both international and national, have focused on studying, developing and demonstrating innovations to this end.[1] The projects aim to enhance the efficiency, speed and reliability of trade and logistics. At the same time, they seek to enhance the effectiveness and efficiency of supervising global trade, safeguarding supply chain security and integrity and safeguarding society against illegitimate trade and criminal threats. Hence, they seek to achieve both business and societal objectives. Balancing the needs for better security with safeguarding economic and public, requires two major transitions:

- Industry and the business communities aim to regain and retain control over fragmented supply chains and need to develop and apply new collaborative ways of supply chain risk management, going beyond the scope of existing enterprise risk management frameworks and building resilience into their supply chains. IT innovations need to support their supply chain visibility and control.
- Governments and border control authorities need to understand and recognise this and develop chain based control and supervision models that are both effective and result in lower compliance costs for legitimate and trusted traders. For governments, this requires a complete paradigm shift from transaction-driven control to system-based control and from a focus on national borders to understanding supply chain dynamics and impact of interventions.

A large part of the solutions currently being developed to support these transitions are IT (-related) innovations. This includes architectures and delivery infrastructures for data sharing in interoperable ways against minimum costs. Innovations are currently being undertaken to enhance data sharing and improve the timely availability of accurate data in global trade networks [3]. This requires business information infrastructures for data sharing. In addition to business-to-business information sharing, global trade networks are highly regulated and for purposes of compliance to e.g. duties and security regulation, companies are required to report a variety of information to various government organisations. This requires intensive information exchange in which a company interacts with, among others, customs authorities. Inter-organisational information systems play a key role in the simplification and harmonisation of cross-border control procedures, where paper-based procedures are replaced by electronic procedures and IT-facilitated inspection and supervision concepts.

The fourth Workshop on *IT-enabled Resilient, Seamless and Secure Global Supply Chains* (WITNESS), organized at the 14th IFIP Conference on e-Business, e-Services and e-Society (I3E 2015), focuses on innovations in the architectures and governance of smart information infrastructures, platforms and ecosystems. Amongst others, the domain deals with the challenges of interconnecting business information infrastructures

[1] For example, European projects in these areas include INTEGRITY, Smart-CM, e-Freight, SUPPORT, LOGSEC, IMCOSEC, CASSANDRA, COMCIS, CONTAIN, SAFEPOST, RISING, ECSIT, ITAIDE, ACXIS, and CORE.

to digital government infrastructures, inter-organizational partnerships, governance and collaboration models. This year's workshop features five papers related to innovations in the architectures of smart digital infrastructures for global supply chains. As an introduction to the workshop, this paper describes recent developments in this domain in terms of an innovation agenda, an overview of the papers, and finalises with a suggestion for additional research topics in need of attention.

2 Making Global Supply Chains Smarter: An Agenda

Among the main innovation drivers in international trade is the need to improve security of international trade supply chains. Security is hampered by a deficient information system, where supply chain actors do not or cannot always fully inform the next actor in the chain, for example about who the original shipper is, and what precisely is in a container [2]. As a consequence, the visibility that for example the consignees have on the goods being shipped to them, is limited. Also, parties further down the supply chain cannot seamlessly integrate goods flows due to a lack of accurate and timely information. What is more, as official declarations are based on transport documents that are not from the actor that really knows the specific goods being shipped, authorities such as customs have to supervise the supply chain with second-hand information that is frequently inaccurate [2].

To make supply chains more secure and to enable parties in the supply chain to enhance supply chain resilience and the effectiveness of operations, innovations that aim to smarten-up international trade are under development. ICT has a big role to play in making supply chains smarter. For instance, innovations in the ICT's used by supply chain actors enable electronic connections and information exchange between the information systems of supply chain actors, thereby enabling or improving access and re-use of these original trade data by other actors in the supply chain [4].

As indicated in the introduction, one place where these innovations are developed and tested is in research and development projects, such as a host of projects funded by the European Commission. As many of these projects are now underway or have finished already, in this section we look at these projects to see what the main innovation categories are to make supply chains more resilient, seamless and secure. We find that three main types of innovations bind these projects:

- Technology innovations;
- New supply chain risk concepts, and;
- New collaborations and supervision concepts.

In the remainder of this section we discuss the main topics that come up in the goals of a selected set of European projects and discuss the abovementioned types of innovations in more detail.

2.1 The Innovation Agenda: Main Topics in Recent EU Projects

There is a series of related EU funded projects that aim to improve the security and economy of international supply chains, by leveraging (information) technology innovations. In this paper, we focus on the following projects: CORE, CASSANDRA,

CONTAIN, SUPPORT, SAFEPOST, EUROSKY, E-FREIGHT, I-CARGO and LOGICON[2], all funded by the European Commission under the 7th Framework Programme. There are many more projects, also outside of Europe and in and by individual countries or stakeholder groups, but we focus on these as they form a 'family' (of sorts) of projects and are funded under the same programme.

A synthesis of the main goals of all aforementioned projects yields the following innovation agenda across these projects:

- Interoperability, information exchange and data sharing;
  - New IT infrastructure and ecosystems (e.g. for air in Europe, for containers globally, for ports);
  - Developing an information backbone; improving security through information-sharing;
- End-to-end supply chain visibility and visibility of supply chain risks;
- Advanced supply chain risk management;
  - Targeted screening (new solutions), improved threat handling;
  - Developing new risk models, from a supply chain perspective;
- System based supervision;
  - A systems approach to design and implement measures for a broader context (e.g. collaborating business in a chain);
  - Risk-based approach to supervision;
- Integration of effective, less-intrusive security technologies in supply chains;
  - Integrating various existing solutions;
  - Container integrated sensors, surveillance system, smart seals and other security technologies;
- Supply chain resilience;
  - Resilience capabilities of organisations and in supply chains;
- Strategic management and governance models;
  - Coordinated border management;
  - Business models for IT infrastructure, supply chain risk management and resilience;
  - Public-private governance models.

Note that optimisations in operations (which has been one of the key topics in supply chains for a long time) are often posited as a derived benefit of these innovations, and not an objective or research component itself, in these projects. Typically, these projects also aim to re-use technologies and innovations that are already present, by selecting the 'best' and integrating them.

An important part of all of these projects is that they aim to demonstrate and refine the innovations in various demonstrators, focusing on container (CASSANDRA), postal (SAFEPOST) and air (EUROSKY) supply chains, including multimodal supply chains

[2] http://www.coreproject.eu/ | http://www.cassandra-project.eu/ | http://www.containproject.com/ | http://www.supportproject.info/ | http://www.safepostproject.eu/ | http://www.euroskyproject.eu/ | http://www.efreightproject.eu/ | http://i-cargo.eu/ | http://www.logicon-project.eu/.

(CONTAIN). From the perspective of the workshop and similar to the CORE agenda, the innovation agenda can be grouped in three main types of innovations, which will be discussed next, based on plans and reports of the projects.

2.2 Technology Innovations

A number of technology innovations act as enablers for the innovation agenda. One of the key enabling technological developments of interest for this workshop are the advancements in ICT. In various ways, the projects mentioned above focus on capturing data from the source, i.e. from the handling actors in activities upstream the supply chain, such as container consolidation and stuffing, and ensuring reuse of high quality data through effective control and validation processes. Most of the papers in this workshop also deal with issues and developments of IT to enable many of the innovation agenda's topics primarily through data sharing. These papers are accompanied by various real demonstrations of such solutions, making the IT innovation an accelerator creating momentum for the other innovations on the agenda.

Reliable source data is a necessary ingredient for realising (amongst others) end-to-end visibility on supply chains and supply chain risk. Such visibility is required to improve company and supply chain performance (by enabling e.g. Vendor Managed Inventory), by supporting and enhancing decision-making processes, agility and resilience. Enhanced visibility promises huge benefits [5], but requires some form of seamless data interoperability and a supporting information infrastructure, for example in the form of an ecosystem (as in e.g. CORE) or a data pipeline (as in e.g. CASSANDRA). The nature of international trade requires a 'federated system-of-systems approach, that connects disparate systems (in supply chains, physical instruments, communities, supply chain security controls) in a global architecture providing effective access to supply chain security related information and services from anywhere within that system' [6].

Seamless data interoperability is aimed at reducing the (transaction) costs involved in searching for and exchanging information and sharing data, in both business-to-business and business-to-government relationships in international trade and logistics. Apart from yielding benefits for businesses, it is expected to contribute to enhanced effectiveness in government control. Lack of interoperability is one of the causes for poor data transformation along the value chain.

Finally, technologies such as detection, scanning and cargo screening technologies, automatic identification and data capture technologies (such as smart container security devices), and tracking and tracing technologies can also contribute to supply chain security and resilience, provided the information generated by all these systems is shared timely and securely.

2.3 New Supply Chain Risk Concepts

Supply chains face various threats and vulnerabilities, ranging from external events to systemic vulnerabilities (such as the information problem discussed before), and now accompanied with concerns about cyber risks, and insurance and trade-finance costs. The visibility that can be provided by IT innovations should also be applied in the portfolio of supply chain risk management options, which is the key to advanced supply

chain risk management. Control over the supply and value chain are hampered by the fragmentation of these chains in terms of actors and interests. Visibility is a key instrument for parties to regain control over the chains they are involved in. Hence, as part of a supply chain wide risk approach and supported by IT innovations, control over and insight in the chains are required for improved alignment of supply chain processes and for supply chain optimisation. However, there is also another economic justification for being in control of the chain; the intrinsic commercial and compliance value of forming a trusted and integrated supply chain. Gartner [7] identifies this as the key trend under Supply Chain Leaders in 2014. Supply chain wide visibility and control over risks across the chain form the foundations for supervision models that are based on the level of business control over the supply chain, which could lead to a lower administrative burden and higher predictability. In any case, a shared risk vocabulary is required, as is improved data and information sharing across supply chain actors (including with government), and building greater agility and flexibility into resilience strategies.

2.4 New Collaboration and Supervision Models

In the EU, DG TAXUD introduced a risk-based approach (RBA) as a customs view on supervision as an alternative to a 100 % scanning approach. Most EU customs organisations already apply an RBA (e.g. by using risk assessment systems). However, these are not always efficient, with up to 30 % of shipments targeted for further evaluation, which results in transaction-based controls disturbing trusted and secure logistics flows. The common framework for Customs Risk Management and Security of the Supply Chain of the European Commission comprises of three parts [8]:

- Identification and control of high-risk goods movements using common risk criteria;
- The contribution of Authorised Economic Operators (AEO) in a customs-trade partnership to securing and facilitating legitimate trade;
- And pre-arrival/pre-departure security risk analysis based on cargo information submitted electronically by traders prior to arrival or departure of goods in/from the EU [8].

Applying such a risk based approach to incoming trade flows highly depends on the quality, accuracy and timeliness of information that supervision agencies receive. From the CASSANDRA project we learned that the data quality of pre-arrival declarations is regularly unsatisfactory. Often, agents or freight forwarders are identified as the consignor and consignee, and customs still does not know who the true consignor and consignee really are. Also cargo related data has quality issues following from these being asked of the ocean carrier, who in many cases does not (even cannot do not not want to) know what is inside a particular container.

The RBAs as developed in the projects also cover what businesses do to identify and deal with risk in their supply chain operations [9]. Specifically, it means that businesses assess the specific risks in their supply chains at the *trade lane level*, and document the controls they have in place to address those risks. Based on that assessment, supply chain risks can be identified that are currently addressed insufficiently by the business controls, and additional controls can be identified to enhance end-to-end control over the supply

chains. This will shift operational focus to higher risk business, thus creating a risk-based approach towards supply chain operations.

Again, the IT innovation plays a major role here, as a supply chain RBA is only possible when the data that circulate among the supply chain partners are accurate, timely and of sufficient quality. Furthermore, the types of risks and the specific sets of data that are required to assess and deal with risks are often product-type and trade lane specific. A supply chain wide risk approach by businesses, combined with a higher quality of crucial data elements, enables government organisations to assess the risks of the supply chain better, and apply a different mix of control mechanisms based on that assessment.

Here, other IT innovations can play a major role as well, primarily in the form of Single Windows [10], which can reduce the administrative burden for businesses, but should also act as a stepping stone for Coordinated Border Management (CBM). CBM refers to a coordinated approach by border control agencies, thus eliminating conflicts and redundancies between different policies, regulations and enforcement practices, thus enhancing their effectiveness as a whole.

An information sharing infrastructure (Sect. 2.2) and novel risk approaches (Sect. 2.3) enable government to 'piggyback' on data from better sources and on business controls; both should be part of a government risk-based approach to supervision; a system based approach (assessing the business controls) and assessing risks based on the data provided to government (for various forms of piggybacking, see [11]). This information must be reliable for government to assess the security of a goods flow, which reduces the risk of unnecessary inspection interference in safe supply chains (but which could not be assessed as being safe because the data is of insufficient quality). As part of this approach, system based control instruments can complement other instruments (such as scanning or physical inspections) in effective control mix, resulting in potentially lower trade transaction costs and lower control burden traders experience.

3 Overview of the Papers in the Workshop

These proceedings contain five papers that are part of the 4th WITNESS workshop. Most of them deal with the challenges of data sharing or with ICT solutions to address them. More specifically, two papers deal with the specifics of data in logistics and international trade, one on the effects of governance of data, the other on assuring data quality. Two other papers deal with the information infrastructure to facilitate data sharing among supply chain actors, including businesses (both traders and logistics providers) and governments (e.g. inspection agencies). Finally, one of the papers deals with a topic that is too often underestimated: whereas the technology might be capable of overcoming many hurdles, ultimately the decision to really adopt and use these technologies depends on many other factors, including the awareness that decision makers have both of problems and of potential solutions.

In the paper "Determining the Effects of Data Governance on the Performance and Compliance of Enterprises in the Logistics and Retail Sector", Nick Martijn, Joris Hulstijn, Mark de Bruijne and Yao-Hua Tan addresses the important topic of

data governance. More specifically, they seek to offer a way to determine the effects that data governance has on business performance and compliance. This is important as many practitioners see or expect positive effects, but are missing a framework to actually assess the effects. This papers offers such a framework and the expected benefits for retail and logistics. The research offers interesting insights that could be added to existing models, including models that start from broader IT governance or enterprise risk management perspectives.

Another paper dealing with the specific characteristics and challenges related to data is "Data Quality Assurance in International Supply Chains: An Application of the Value Cycle Approach". In this paper, Yuxin Wang, Joris Hulstijn and Yao-Hua Tan propose a value cycle approach to assuring data quality in international supply chains. Given the numerous issues existing when it comes to data quality in international trade, the insights that the paper offers might be helpful to academics and practitioners looking into ways to ensure and verify the quality of the data that supply chain actors have to work and make decisions with.

In his paper "Towards a federated infrastructure for the global data pipeline", Wout Hofman offers set of platform services and protocols to allow interoperability of different platforms. Such interoperability is necessary given that global data exchange infrastructures (such as the data pipeline that this paper concerns) will have to be federated solutions, making use of the wide variety of solutions already in place or currently under development. The path to the vision laid out in this paper will not only require time and continued innovation, but will also be in need of new business models to ensure adoption and sustainability of the federated infrastructure.

Thomas Jensen also discusses infrastructure for improving information sharing in international trade, using the Interorganisational Systems (IOS) literature as his starting point. In "Key Design Properties for Shipping Information Pipeline", he follows a design science approach and presents a set of key design properties for a so-called Shipping Information Pipeline, an information infrastructure for international containerized trade. Not unlike Hofman's ideas, one of the principles also holds that the idea of the data pipeline is that of a 'virtual' infrastructure, whereas the actual physical infrastructure can (and will) be handled by several individual organisations.

Finally, in "Enhancing Awareness on the Benefits of Supply chain Visibility through Serious Gaming", Tijmen Joppe Muller, Rainer Müller, Katja Zedel, Gerwin Zomer and Marcus Engler discuss a serious game developed for customs-related issues and innovations in international supply chains. The gaming-background of the paper primarily concerns the learning effect, but given the background of the innovation that it concerns (data pipeline, customs innovation), we think that it can also serve other purposes as part of the innovation trajectory, notably by supporting the process of activating and including key stakeholders. The main contribution of this paper is therefore twofold: it discusses the specific game, which covers various topics of the innovation agenda discussed in this paper, but it also shows how serious gaming could (and should) be used to address non-technical aspects of the innovations.

4 Beyond the Innovation Agenda: Making Innovations Count

As demonstrated by the innovation agenda of the projects and the papers in this workshop, logistics and international are seeing a series of innovations to make logistics smarter, more competitive, improve supply chain security, enhance control and facilitate information sharing. These innovations have been and are being developed, studied and demonstrated extensively in (inter)national and European projects and are about new ways of controlling risk and collaborating in supply chains, enabled by IT innovations. Combined, these innovations should constitute a transition for the domain of logistics and government. Although the concepts that are central to these innovations are quite well known by now (as demonstrated by the many shared goals of the projects covered in this paper), there still is a big gap between the transition in theory and the actual adoption thereof in the real world. The real uptake and in-depth integration in the actual practices of organisations of what has come from these years of research, concept building, and technical development, is unfortunately limited at times. That some concepts only find limited fertile ground in the sector after the projects finish, presents a major challenge for the coming years. The limited adoption of the ideas and innovations generated by the research projects itself represents a research problem, which cannot be solved only by improving the models and technologies.

Although the European projects devote a lot of time and resources on stakeholder-driven demonstrators (which is a very important step), just how these local innovations combined can constitute a coordinated and balanced innovation for the logistics sector as a whole, is not yet well understood. A major transition through an integrated and coordinated large-scale implementation of key innovative logistics concepts is needed, but the actual adoption by the sector also requires small, local innovations, attuned to the specific stakeholders involved.

Hence, although further refinement and additional development of the topics on the innovation agenda discussed in this paper are both needed and desired, we argue that additional research is needed into the *socio* side of what ultimately are to be socio-technical artefacts. This side has challenges itself, e.g. as the parties involved in logistics and trade, which ultimately will need to adopt the innovations and solutions, have great variety in position, interests and values. To make the innovations work for them, this variety will have to be addressed and taken into account, which can only be done close to those specific stakeholders. This is difficult as the stakeholder field is very fragmented, with interests divvied up along the lines of sectors (e.g. initiatives in the horticulture industry), roles in the supply chain (e.g. transporters, freight forwarders, shippers) and the size of companies. Especially the many small and medium sized operators will have to be involved via e.g. branch organisations and collaborations as they have limited capacity focused on innovation. This complexity can explain, for example, why some of the aforementioned projects have become innovations that rely on government support, whereas they should have been (and in nature are) business innovations.

To create the right incentives for open innovation and establish a vibrant community of companies that are willing provide parts for data pipelines and to ensure that public value is realized, institutional arrangements must be developed and adopted. Typically, these institutional arrangements are developed not only by national government, but also

supra-national organizations should be involved such as the European Commission or the United Nations. Furthermore, as argued by Hofman in his workshop paper, apart from the IT developments and challenges, the federation of platforms requires additional research into the (sustainability of) business models for these platforms. Making combinations of data from multiple parties (both public and private) is essential for developing new commercial services and for supporting new supervision concepts, both creating (economic) incentives for companies to contribute to realizing the innovations. However, there are also risks involved in combining data from multiple sources, especially in a competitive private environment. Some form of governance would help stakeholders (again, both public and private) to create on the one hand a level playing field (e.g. to avoid that one or a few parties can gain an unfair competitive advantage from their access to community data), and on the other hand offers enough economic incentives for businesses to make their adoption of the innovations commercially viable. As the workshop paper by Muller et al. illustrates, much of this starts with making key players aware of the issues and potential solutions. Given such awareness, the right institutional practices, incentives, process support by facilitators and viable business models for parts of the innovations, the field can develop a fertile ground for the innovations to land in, and from thereon find wide support and uptake.

Acknowledgements. This work is part of the research project "Governing public-private information infrastructures", which is financed by the Netherlands Organisation for Scientific Research (NWO) as Veni grant 451-13-020. Furthermore, for this paper, we made use of parts of CASSANDRA and CORE project reports, both funded by the European Commission as part of the 7th Framework Programme (FP7). The authors are involved in both.

References

1. Van Oosterhout, M.: Organizations and flows in the network. In: Van Baalen, P., Zuidwijk, R., Van Nunen, J. (eds.) Port Inter-Organizational Information Systems: Capabilities to Service Global Supply Chains, vol. 2, pp. 176–185. Now Publishers, Hanover (2009)
2. Hesketh, D.: Weaknesses in the supply chain: who packed the box? World Customs J. **4**, 3–20 (2010)
3. Klievink, B., Van Stijn, E., Hesketh, D., Aldewereld, H., Overbeek, S., Heijmann, F., Tan, Y.-H.: Enhancing visibility in international supply chains: the data pipeline concept. Int. J. Electron. Gov. Res. **8**, 14–33 (2012)
4. Tan, Y.-H., Bjørn-Andersen, N., Klein, S., Rukanova, B. (eds.): Accelerating Global Supply Chains with IT-Innovation. ITAIDE Tools and Methods. Springer, Berlin (2011)
5. Ngai, E.W.T., Chau, D.C.K., Chan, T.L.A.: Information technology, operational, and management competencies for supply chain agility: findings from case studies. J. Strateg. Inf. Syst. **20**, 232–249 (2011)
6. CORE Consortium: CORE (Consistently Optimised Resilient Secure Global Supply-Chains) Description of Work (DoW) (2014)
7. Gartner: The gartner supply chain top 25 for 2014 (2014)
8. European Commission: Communication from the commission to the european parliament, the council and the european economic and social committee on Customs Risk Management and Security of the Supply Chain (2013)

9. Cassandra: D2.2 – risk based approach (2012)
10. Keretho, S., Pikart, M.: Trends for collaboration in international trade: building a common Single Window Environment. United Nations (2013)
11. Klievink, B., Bharosa, N., Tan, Y.H.: Exploring barriers and stepping stones for system based monitoring: insights from global supply chains. In: Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI), pp. 35–42 (2013)

Determining the Effects of Data Governance on the Performance and Compliance of Enterprises in the Logistics and Retail Sector

Nick Martijn(✉), Joris Hulstijn, Mark de Bruijne, and Yao-Hua Tan

Faculty of Technology, Policy and Management,
Delft University of Technology, Delft, The Netherlands
nickmartijn@gmail.com,
{j.hulstijn,m.l.c.deBruijne,y.tan}@tudelft.nl

Abstract. In many of today's enterprises, data management and data quality are poor. Over the last few years, a new solution strategy has emerged, known as data governance: an overarching methodology that defines who is responsible for what data at what point in a business process. Although positive effects on the business performance and compliance of enterprises are seen in practice, a substantiated method for determining the effects of data governance has not yet been developed. This paper reports on explorative research to develop such a specification method. Through a conceptualization of data governance based on literature, case study analysis of clients of a large consultancy firm and interviews with representatives of companies that have recently implemented data governance, an effect specification framework was developed. Using the interviews, initial steps towards validation were performed.

Keywords: Data governance · Effect specification

1 Introduction

Enterprise data is becoming increasingly important. Data was initially seen as a byproduct of business processes, used for example for financial recording (Lake and Crowther 2013). Nowadays, data is considered to be a valuable asset in and of itself (Bughin et al. 2010). This value is primarily provided by two applications: measuring business performance and compliance reporting. Firstly, increasingly complex and globalizing business processes require the support of reliable data. For example, the international container shipping industry requires timely and accurate data to feed its logistical planning. Lack of data quality leads to huge losses (Steinfield et al. 2011). Secondly, enterprise data is used for financial reporting and for other kinds of compliance reporting. Companies have to comply

This paper summarizes (Martijn 2014). The research has been conducted with the support of Marinka Voorhout, specialist in Enterprise Data Management. We gratefully acknowledge her contribution.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 454–466, 2015.
DOI: 10.1007/978-3-319-25013-7_37

to certain laws, such as Sarbanes-Oxley for companies listed on the US stock exchange, or Solvency II for insurance companies (Eling et al. 2007). These laws demand that companies demonstrate to the regulator that they are compliant, which requires evidence. Regulatory compliance creates additional data requirements. It is not sufficient to supply evidence; an audit trail is also required (Jiang and Cao 2011). Not meeting data requirements can lead to severe financial consequences. Consider the $3.75 m fine Barclays bank received from US Financial Industry Regulatory Authority (BBC 2013). So enterprise data is used to gain insight in business performance, while also enabling compliance (Cheong Chang 2007; Golfarelli et al. 2004; Loshin 2012).

Notwithstanding its importance, the standard of data management is often poor (Haug et al. 2011). Although most companies have well-managed IT systems, the responsibilities for maintaining specific kinds of data are mostly not incorporated (Redman 2001). It is has been shown that when no clear policies, rules and controls are defined within the organization about who is responsible for what data, overall data quality will deteriorate (Batini et al. 2009). Poorly governed data may generate losses, as incomplete or erroneous information can mean a serious strategic disadvantage or lead to inefficiently organized business processes (Steinfield et al. 2011), in addition to posing the risk of being deemed non-compliant.

Over the last few years a new solution strategy has emerged: *"Data governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods"* (Thomas 2006). Essentially, data governance is an overarching methodology that defines who is responsible for what data at which point in the process. There is more to it though, such as internal controls, information systems architecture, standardization of data formats, corporate culture and use of technology, such as monitoring tools. Taken together, data governance measures can assure that enterprise data will be of sufficient quality. Data quality is seen as the most important aspect influencing usability of data for business processes and reporting (Friedman and Smith 2011).

It turns out to be relatively hard to specify the effects of data governance projects and interventions. How do the application of various tools and techniques affect data quality? And subsequently, how does improved data quality affect business performance and compliance reporting? These are fundamental questions that have received relatively little attention. The effectiveness of data governance projects is only known from practical experience (De Waal and De Jonge 2012). Data governance frameworks, such as DAMA DMBOK (Mosley et al. 2010), claim that they can determine these effects, but these are not fully scientifically substantiated.

In this paper, we therefore develop a framework to make it possible to specify the effects of data governance interventions. The paper is a summary of the graduation research reported in (Martijn 2014). We make use of case studies of firms, which have recently undertaken data governance projects. These cases were collected with the help of a large consultancy firm with extensive experience in helping clients improve their data quality through a data governance framework, in particular in the financial, logistics and retail sector (De Waal and De Jonge 2012).

Concerning the choice of research method, note that organizational factors may affect the effectiveness of data governance interventions, but cannot be unambiguously operationalized. Thus, the case study approach is most appropriate, as the boundaries between the phenomenon (data governance) and the context (organizational effects) is relatively unclear (Boschi 1982; Xiao et al. 2009).

The research proceeds as follows. Based on literature, we develop a conceptualization of data governance and its drivers (Sect. 2). We then develop a method to specify the expected effects of data governance interventions (Sect. 3). The conceptualization is relatively generic: it must be further specified for each case. Making use of client dossiers of companies that have recently adopted data governance measures, we show how the concepts can be further operationalized. Based on interviews with representatives of companies from the retail and logistics sector that are currently implementing data governance measures, we take initial steps towards validation of (Sect. 4). Full validation would require more cases, and would require comparison of the outcomes with other, independent, specification techniques.

2 Conceptualizing Data Governance

There is a lot of research on data quality and the effect it has on the use of information systems (Strong et al. 1997). However, not much scientific research is specifically dedicated to the reverse question. How does data governance improve data quality, and consequently increase business performance and compliance? There are several data governance frameworks, of which the DAMA Data Management Body of Knowledge is most commonly used (Mosley et al. 2010). Such frameworks provide an overview of data governance measures to increase the data quality at an organization. The DAMA approach summarizes the following best practices (Mosley et al. 2010): data architecture management, data development, database operations management, data security management, reference and master data management, data warehousing and business intelligence management, document and content management, meta data management, data quality management, all centered around the data governance. A problem with such frameworks is that there are generic and professionals have to adapt them to their own situation. The framework lists different kinds of activities, both technical and strategic. How can we structure their dependencies?

We have made a conceptualization of data governance, shown in Fig. 1. As argued in the introduction, data quality is an essential property driving business performance and compliance. This will therefore be used as the guiding notion. To structure the diagram and locate the various activities, we use an enterprise architecture, based on the layers of (Winter and Fischer 2006). From the bottom up: technology, software and integration architecture (merged here), process architecture, and business architecture. On the right we added organizational architecture, as we focus on governance aspects, involving roles, responsibilities and institutional arrangements.

Reviewing the structure from the bottom up, the first part consists of the technology, software and integration architecture. Here we find physical devices (gates; RFID readers etc.), computer systems (databases; networks etc.) and software applications

(ERP systems; workflow management etc.) to store, retrieve and process information. In addition, we also find protocols and procedures for exchanging information and for integrating different modules. In this layer the actual data is situated.

Above this infrastructure layer, the process architecture is located, including the processes that are carried out with the data. Data processes are composed of four basic operations: Create, Read, Update and Delete (Martin 1983; Polo et al. 2001). These CRUD operations determine the status of data elements at any point in the process, so we could say that this layer also contains the data architecture, which determines how data is being handled. When a certain piece of data, such as delivery address or price of a product is used, the data is mostly pulled from the infrastructure through an Enterprise Resource Planning (ERP) system.

The business architecture layer includes the value-adding processes, such as purchasing, sales, manufacturing or transport. Also internal control and risk management, compliance management, and reporting (financial statements, tax reports) are located here. To enable successful business processes, an effective data and process architecture is required. Consider for example the process of acquiring resources from a supplier, manufacturing products, and selling them on to a customer, consisting of steps like: receive, pay, manufacture, store, sell, dispatch. These steps in the primary process give rise to data operations. For example, receiving resources in a warehouse means the creation of new data objects representing the type of resources, storing the stock levels for these resources, updating inventory in the general ledger, changing the status of the corresponding purchase order, etc. The way in which basic 'CRUD' operations are implemented, largely determines data quality (Wand and Wang 1996).

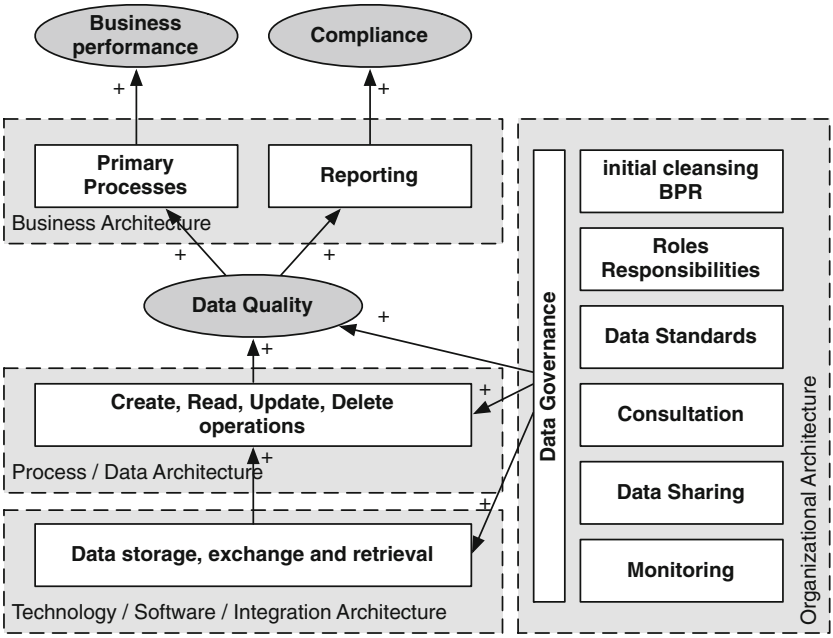


Fig. 1. Conceptualization of data governance and its drivers

Kahn et al. (2002) see data quality as the degree to which data is fit for purpose, i.e. meeting company-specific requirements, see also Juran et al. 1999 and Wang et al. (1996). Data quality in turn largely determines the effectiveness of the business processes, and also influences the reporting quality (Eppler and Helfert 2004), thereby affecting the drivers: business performance and regulatory compliance. Low data quality is pervasive, costly and can cause high inefficiencies (Eppler and Helfert 2004; Fisher and Kingma 2001; Wang et al. 2001).

The organizational architecture contains institutional bodies like the board of directors, managing staff, change advisory board (CAB) and so forth. As data governance is a top-down methodology, we also locate most of the data governance efforts here. In particular, it entails roles for making people responsible for enterprise data. Data governance can influence all other layers. There is also a direct influence link between data governance and data quality, because of initial cleansing activities that are often part of a data governance project (see Sect. 2.1).

Business performance and regulatory compliance are at the end of the causal chain. First, the data that is used in business processes, directly affects business performance (Neely et al. 2002). Errors lead to missed deliveries, dissatisfied customers, etc. Also internal reports are used for forecasting, budgeting etc. Second, the reporting function influences regulatory compliance, as evidence needs to be produced of being compliant with laws and regulations (Jiang and Cao 2011).

2.1 Elements of Data Governance

Given this conceptualization of data governance, what does it actually involve? Data governance consists of various interrelated elements:

- *Initial cleansing* and *business process redesign* interventions are required before data governance can function properly within an organization. This sets a basic data quality level at the start of a data governance project. Often it also involves redesign of CRUD processes, compare process redesign (Hammer 1990).
- *Roles and responsibilities* are essential to prevent lack of clear ownership for data management. When someone is made responsible, it can be assumed that less errors will enter the system, and that errors are detected and solved earlier, leading to more efficient processes (Mosley et al. 2010).
- *Data standards* describe how to represent, process, use and handle enterprise data. Implementing data standards, in combination with a governance structure that enforces the standards, leads to higher data and process quality. Use of data standards is a prerequisite for other data governance interventions. Standards also make it possible to measure data quality levels to indicate progress.
- C*onsultation* is meant to improve communication between departments (horizontal), and between management levels (vertical). Besides communication enhancements in primary processes and workflows, so called consultation platforms are recommended to improve the adaptability of data governance measures themselves. Errors should be identified and traced by feedback from users (Orr 1988), so the company can learn from experience.

- *Data sharing* with supply chain partners is important for efficient alignment within supply chains, both internally, as well as externally (Steinfield et al. 2011). This includes monitoring of data provided by the supplier to ensure sufficient quality. Data protocols can be part of the contract provisions.
- *Monitoring* should provide continuous insight in the current quality of the data to facilitate manageability of data by responsible employees, for instance, implementing tools that produce real-time data quality overviews on a dashboard. Again, this requires the ability to measure data quality level.

This list shows that data governance intervention requires a form of governance: it cuts across all layers and departments, which requires management support. Business must be involved, as they should define the information needs. Standards must be enforced. Budget to make the required changes to the IT infrastructure must be secured. Furthermore, even if we narrowly define data governance as the implementation of roles and responsibilities over enterprise data (Thomas 2006), it cannot be abstracted from other data management aspects such as the use of standards and tools. After all, the roles and responsibilities are meaningless without technical and organizational means to support employees in executing these responsibilities. Therefore data governance is seen as a 'package deal': these elements strengthen each other.

3 Deriving a Causal Model

Using the diagram in Fig. 1, we developed a causal model to specify the effects of data governance on an organization, shown in Fig. 2. As such models are typically domain specific, the contribution lies in the method to derive the model. The model is developed on the basis of scientific literature and case study research (Sect. 4). Insights are based on dossier reviews at a large consultancy firm and interviews with clients who have recently been advised on data governance. Versions of the model were validated and adjusted on the basis of interviews with clients.

To scope the research, we decided to focus on cases from the logistics and retail sector. Data quality within supply chains is highly important (Li and Lin 2006). Business performance can be operationalized using key performance indicators from the Supply Chain Operation Reference (SCOR) model. This model provides a standard method to review the performance of a supply chain, see Lockamy and McCormack (2004), Xiao et al. (2009) and Hwang et al. (2008). The KPIs include seven elements to determine customer service level: right product, right customer, right time, right place, right condition, right quantity and right costs (Fawcett and Fawcett 2014).

The literature research, project dossiers and interviews with experts produced hypotheses for relationships between data governance measures, data quality, and ultimately business performance and compliance. The relations are shown as arrows in Fig. 2. We use the following semantics: A –[+]– > B means a positive influence: when A increases, B should also increase. Conversely, A –[−]– > B means a negative influence. When A increases, B should decrease.

Starting at the left part of Fig. 2, the various data governance elements will lead to better CRUD operations in business processes and subsequently, to better data quality. Initial cleansing will improve data quality directly. Consultation will improve

communication between departments, which may lead to better IT responsiveness: the ability of the IT department to meet business demands. The enforcement of standards will reduce introduction of mistakes; in addition, it will make it easier to measure data quality. After all data quality is defined as fitness for purpose, where the purpose is reflected in company policies and data requirements, such as for instance those suggested by the SCOR model. Improved data sharing between partners in the supply chain, will improve data quality from suppliers. So also external factors play a role.

In the middle, better data quality helps to improve supply chain forecasting quality (Shankaranarayanan and Cai 2006). Inherently, if forecasts are not reliable, the primary processes will be run less efficiently (Gunasekaran et al. 2004).

Improved data quality also decreases administrative costs. According to a worldwide investigation by (GS1 2011), low data quality causes significant administrative costs referred to as shrinkage, the difference between what is shipped by the supplier and what is finally sold to the customer. Furthermore, efficiency in the primary process reduces operating expenditure (OPEX). Operating expenditure consists of all costs associated with operating a supply chain, such as transport and transaction costs. In particular, administrative costs have a large impact on OPEX.

When primary processes are more effective and efficient, for example when timeliness of deliveries is increased, the level of customer service will increase (Stevenson and Hojati 2007). Customer service crucially affects sales. In addition, customer responsiveness is defined as the manner in which the business can meet demands of customers (Friedman and Smith 2011). This property is related to the infrastructure: can it adapt. If a number of basic data elements are collected and processed reliably, new combinations can be engineered relatively easily. This improves the ability to forecast and report but also the ability to construct new customer services. These factors will help to increase sales and thus business performance (Neely et al. 2002).

In the lower part of the diagram, higher data quality leads to improved internal controls and reporting quality, which by definition increases regulatory compliance. In most cases, an audit trail of enterprise processes and cash flow is required. Reporting quality is lower, when material (i.e. crucial) errors are not detected, or when relevant aspects of behavior are not reported, or not even recorded. Data quality is closely related to the notion of reliability, which involves accuracy (data correspond to reality) and completeness (all relevant aspects of reality are recorded) (Strong et al. 1997).

In addition, reporting quality is affected by regulatory responsiveness, the ability to deal with compliance demands (Friedman and Smith 2011). Data governance also affects this responsiveness variable: when basic figures are recorded and processed reliably, with an audit trial, new combinations of reports can be constructed reliably.

The relation between compliance reporting and data quality becomes even more crucial, when we apply innovative ideas of regulatory supervision, in which data is pulled from the source. For example, a ‘data pipeline’ infrastructure could facilitate reliable exchange of information in a trade lane, with access for customs authorities, but also for authorized traders (Klievink et al. 2012). Also in the XBRL-GL vision, data items are recorded and ‘tagged’ close to the source (Cohen 2009). This makes it possible to record an audit trail with meta-data about provenance of data items. Given such basic elements with their provenance, new reports can be constructed reliably.

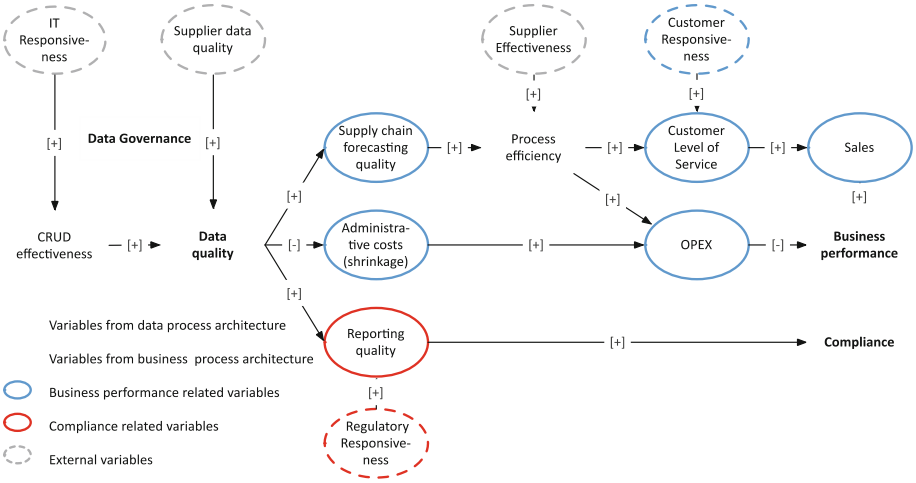


Fig. 2. Causal diagram: expected effects of data governance

The discussion above shows, that it is in fact possible to operationalize the effects of data governance. Initially this will be qualitative, but once the organization has gained some experience also quantitative measures can be used. The resulting model always depends on the specific case; what matters is the line of reasoning. We identify variables that can be used to monitor the effects of data governance interventions.

Some of these variables are well defined and are measured on a routine basis: OPEX, sales, customer service level. Other variables can be defined, once a good operationalization is found. CRUD effectiveness can be determined by a process review, as is typically done by an operational auditor. For instance, consider the property that all information needs should be covered, or that no data should be requested that is not used later. Such properties can be determined using a create-use matrix, in which all data elements are plotted against the activities and roles by which they are created, and subsequently used. Data quality can be specified as the converse of the number of deviations per volume of data from a set of agreed data requirements. So we have *DQ = volume of data/number of deviations*. In the supply chain domain, it makes sense to use the SCOR indicators as a starting point for those requirements. The point is to trace supply chain delivery errors back to the information systems that cause them. Supply chain forecasting quality is determined by comparing forecasts with actual performance. Reporting quality is typically determined as a by-product of a financial audit: the number of deviations is reported as well. The variables IT responsiveness, customer responsiveness and regulatory responsiveness should be seen as intermediate variables, to factor in the ability of the organization to adapt. Typically, these can't be measured, which is why they are indicated as dashed ellipses.

That leaves only process efficiency, business performance, and compliance. Those variables can be seen as outcomes of data governance interventions, not inputs that can be used to control and adjust. These outcomes can in fact be measured, but that topic is

out of scope. Consider for instance the Business Balanced Scorecard to measure business performance. Here we focus on the effects of data governance.

Table 1. Overview of cases and main impact on the business

	Market	Main characteristic	Impact on the business
A	Food retail	Insufficient data quality in the product database, primarily caused by lack of formalized responsibilities	The company was able to improve supply chain efficiency and decrease operational expenditure. For example, 50 % of working hours on data related activities was saved
B	Food production	Strong focus on the centralization and standardization of data management	Costs of production could be lowered. Business strategy improved due to improved data quality and better insight in enterprise data
C	Logistics	Strong focus on data quality due to a supply chain in a competitive market and compliance issues	Decision making and efficiency within the supply chain improved significantly, leading to fewer compliance issues and higher customer satisfaction
D	Financial sector	Customer data stored in many different legacy systems, leading to inconsistent, incomplete, incorrect and double entries	Improved data quality increased client satisfaction and improved regulatory compliance

4 Towards Validation

To assess the adequacy and usefulness of the causal model, representatives of four companies that have recently implemented data governance were interviewed. The interviews were conducted at companies from the retail and logistics sector, which aligns with the research scope. For comparison, one case from the financial sector was included. We found no structural differences in responses from the domains.

The interviewees are mostly (IT) managers that were closely involved in the implementation of data governance measures, based on advice from the consultancy firm. Generally, their overall opinion on data governance is highly positive. They see the added value of data governance and acknowledge the positive effect on data quality and therefore on compliance and business performance. Issues identified are mostly on an operational level (Table 1). An example is erroneous product information leading to inefficiencies due to data errors (case A).

Firstly, representatives of all four companies confirmed the presence of the main causal relations in the framework. They too experienced that data governance measures have an effect on the implementation of CRUD processes, on the data quality level, and on the primary processes. This can be seen as a first step towards validation of the methodology that was developed in this research. Secondly, based on these interview outcomes, the early findings from the project dossier research and literature research, summarized in Fig. 1, can also be validated. Thirdly, the interviews led to the observation that the further 'downstream' a factor is in the framework, the more difficult it is to specify precisely. It could be that this originates from the fact that most interviewees had a strong 'data-view' on the business, in which CRUD processes and data quality play an important role. For most of the interviewees, business issues are out of their scope. Fourthly, it is confirmed that data governance is a top-down methodology, as was found in the literature research. Measures taken to implement data governance, such as standards and tooling (monitoring) are forced onto the company by management responsible for data governance, backed by general management. This result supports the organizational theory behind the conceptualization.

5 Conclusions

Business performance measurement and compliance reporting are driving an increasing demand for improved data quality in enterprise systems. Data governance has emerged as a solution concept. It is hard to specify the intended effects of a data governance project before the start of the project: what is the business case? It is even harder to specify the actual effects afterwards, both qualitatively, i.e., has data quality improved; has client market responsiveness improved, and quantitatively, i.e., how much has business performance improved; how much costs have been saved?

In this research we have studied literature about data governance, both from theory and practice. We have reviewed dossiers of recent data governance projects conducted at a consultancy firm, and have held interviews with experts and with clients of this firm. This material has led to two outcomes:

(1) a conceptualization of data governance, positioning it within the organizational part of an enterprise architecture, and indicating the effects on its drivers, namely business performance and regulatory compliance.
(2) a causal model, with hypotheses about the influence of data governance on variables representing data quality and other intermediate notions, and ultimately on business performance and compliance.

Precisely specifying the effects of data governance interventions is exceptionally complex. Data governance is an overarching and top-down methodology. Many sector-specific factors are involved. Moreover, many of the factors cannot be operationalized. For instance, every measurement of data quality – correspondence to data requirements that represent fitness for purpose – depends on the definition that is applied by a specific enterprise. The data requirements, standards or data rules are

always changing, because businesses are dynamic and respond to market conditions. Furthermore, in terms of effect determination, data governance cannot be abstracted from other data management practices. Data governance involves a package deal, of measures that mutually strengthen each other. All this makes a generic specification of the effects of data governance impossible.

We can however provide a method of how data governance effects can be specified within a sector (using e.g. the SCOR model in supply chain management), or within a specific enterprise. The causal model in Fig. 2 can serve as a 'back bone' for such a method. It follows the logic of increasing positive effects (customer satisfaction; sales) and reducing negative effects (operational expenditure), by reducing data errors and improving process efficiency, as well as reporting quality.

The set-up of the research certainly has limitations. First, the research is focused on the logistics and retail sector. When another sector is considered, only the main relations in Fig. 2 can be used, not the choice of measures. Second, because data governance is an overarching methodology, in which measures strengthen each other, it is impossible to study the individual effectiveness of interventions. Such evaluations would be valuable in practice to improve efficiency of projects. Thirdly, the research is scoped towards the expected benefits of data governance. Especially for business purposes, in which benefits are often weighed against costs, it would be useful to gain insight in the costs in order to build a proper business case for data governance projects. Consider IT investments, costs of additional personnel, costs of maintenance of tooling, and increased controls. These investments should be balanced against the costs of poor data governance. This is a useful topic for future research.

References

Batini, C., Cappiello, C., Francalanci, C., Maurino, A.: Methodologies for data quality assessment and improvement. ACM Comput. Surv. (CSUR) **41**(3), 16 (2009)

Boschi, R.: Modelling exploratory research. Eur. J. Oper. Res. **10**(3), 250–259 (1982)

Bughin, J., Chui, M., Manyika, J.: Clouds, big data, and smart assets: ten tech-enabled business trends to watch. McKinsey Q. **56**(1), 75–86 (2010)

BBC: Barclays fined $3.75m after record-keeping failure, from BBC News (2013). http://www.bbc.co.uk/news/business-25525621

Cheong, L.K., Chang, V.: The need for data governance: a case study (2007)

Cohen, E.: XBRL's global ledger framework. Int. J. Discl. Gov. **6**(2), 188–206 (2009)

De Waal, A., De Jonge, A.W.A.: Data Goverance bij een grote verzekeraar. Compact **39**(2), 11–17 (2012)

Eling, M., Schmeiser, H., Schmit, J.T.: The solvency II process: overview and critical analysis. Risk Manage. Insur. Rev. **10**(1), 69–85 (2007)

Eppler, M., Helfert, M.: A classification and analysis of data quality costs. In: Paper presented at the International Conference on Information Quality (2004)

Fawcett, S.E., Fawcett, A.M.: The Definitive Guide to Order Fulfillment and Customer Service: Principles and Strategies for Planning, Organizing, and Managing Fulfillment and Service Operations. Pearson Education Inc., Upper Saddle River (2014)

Fisher, C.W., Kingma, B.R.: Criticality of data quality as exemplified in two disasters. Inf. Manage. **39**(2), 109–116 (2001)

Friedman, T., Smith, M.: Measuring the Business Value of Data Quality. Gartner, Stamford (2011)

Golfarelli, M., Rizzi, S., Cella, I.: Beyond data warehousing: what's next in business intelligence? In: Paper Presented at the Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP (2004)

GS1. Australia Data Crunch Report. GS1, Sydney, Australia (2011)

Gunasekaran, A., Patel, C., McGaughey, R.E.: A framework for supply chain performance measurement. Int. J. Prod. Econ. **87**(3), 333–347 (2004)

Hammer, M.: Reengineering work: don't automate, obliterate. Harvard Bus. Rev. 104–112 July-August 1990

Haug, A., Zachariassen, F., Van Liempd, D.: The costs of poor data quality. J. Ind. Eng. Manage. **4**(2), 168–193 (2011)

Hwang, Y.-D., Lin, Y.-C., Lyu Jr., J.: The performance evaluation of SCOR sourcing process—the case study of Taiwan's TFT-LCD industry. Int. J. Prod. Econ. **115**(2), 411–423 (2008)

Jiang, K., Cao, X.: Design and implementation of an audit trail in compliance with US regulations. Clin. Trials **8**(5), 624–633 (2011)

Juran, J.M., Godfrey, A.B., Hoogstoel, R.E., Schilling, E.G.: Juran's Quality Handbook, vol. 2. McGraw Hill, New York (1999)

Kahn, B.K., Strong, D.M., Wang, R.Y.: Information quality benchmarks: product and service performance. Commun. ACM **45**(4), 184–192 (2002)

Klievink, B., Van Stijn, E., Hesketh, D., Aldewereld, H., Overbeek, S., Heijmann, F., Tan, Y.-H.: Enhancing visibility in international supply chains: the data pipeline concept. Int. J. Electron. Gov. Res. **8**(4), 14–33 (2012)

Lake, P., Crowther, P.: Concise Guide to Databases. Springer, London (2013)

Li, S., Lin, B.: Accessing information sharing and information quality in supply chain management. Decis. Support Syst. **42**(3), 1641–1656 (2006)

Lockamy, A., McCormack, K.: Linking SCOR planning practices to supply chain performance: an exploratory study. Int. J. Oper. Prod. Manage. **24**(12), 1192–1218 (2004)

Loshin, D.: Evaluating the Business Impacts of Poor Data Quality. Knowledge Integrity Inc, Silver Spring (2012)

Martijn, N.: Exploring the Effects of Data Governance. Msc thesis, Delft University of Technology, Faculty of Technology, Policy and Management (2014)

Martin, J.: Managing the Data Base Environment. Prentice Hall, Upper Saddle River (1983)

Mosley, M., Henderson, D., Brackett, M.H., Earley, S.: DAMA guide to the data management body of knowledge (DAMA-DMBOK guide). Technics Publications (2010)

Neely, A.D., Adams, C., Kennerley, M.: The Performance Prism: the Scorecard for Measuring and Managing Business Success. Prentice Hall Financial Times, London (2002)

Orr, K.: Data quality and systems theory. Commun. ACM **41**(2), 66–71 (1988)

Polo, M., Piattini, M., Ruiz, F. Reflective persistence (Reflective CRUD: reflective create, read, update and delete). In: Paper Presented at the Sixth European Conference on Pattern Languages of Programs (EuroPLOP) (2001)

Redman, T.C.: Data Quality: the Field Guide. Butterworth-Heinemann, Wobum (2001)

Shankaranarayanan, G., Cai, Y.: Supporting data quality management in decision-making. Decis. Support Syst. **42**(1), 302–317 (2006)

Steinfield, C., Markus, M.L., Wigand, R.T.: Through a glass clearly: standards, architecture, and process transparency in global supply chains. J. Manage. Inf. Syst. **28**(2), 75–108 (2011)

Stevenson, W.J., Hojati, M.: Operations Management, vol. 8. McGraw-Hill/Irwin, Boston (2007)

Strong, D.M., Lee, Y.W., Wang, R.Y.: Data quality in context. Commun. ACM **40**(5), 103–110 (1997)

Thomas, G.: Alpha Males and Data Disasters: the Case for Data Governance. Brass Cannon Press (2006)

Wand, Y., Wang, R.Y.: Anchoring data quality dimensions in ontological foundations. Commun. ACM **39**(11), 86–95 (1996)

Wang, R.Y., Mostapha, Z., Yang, W.L.: Data Quality. Kluwer Academic Publishers, Dordrecht (2001)

Wang, R.Y., Strong, D.M., Guarascio, L.M.: Beyond accuracy: what data quality means to data consumers. J. Manage. Inf. Syst. **12**(4), 5–33 (1996)

Winter, R., Fischer, R.: Essential layers, artifacts, and dependencies of enterprise architecture. In: EDOCW 2006 (2006)

Xiao, R., Cai, Z., Zhang, X.: An optimization approach to cycle quality network chain based on improved SCOR model. Prog. Nat. Sci. **19**(7), 881–890 (2009)

Data Quality Assurance in International Supply Chains: An Application of the Value Cycle Approach

Yuxin Wang(✉), Joris Hulstijn, and Yao-Hua Tan

Section Information and Communication Technology (ICT), Faculty of Technology, Policy and Management (TBM), Delft University of Technology, Delft, The Netherlands
Y.Wang-12@tudelft.nl

Abstract. With increasing international trade and growing emphasis on security and efficiency, enhanced information and data sharing between different stakeholders in global supply chains is required. Currently data quality is not only problematic for traders, but also for various government agencies involved in border control, such as customs authorities and border force. We adapt principles from value cycle modelling in accounting, and show how these principles enabled by ICT can be extended to supply chain management to ensure quality of data reported to customs. We then describe a typical application scenario based on a real but anonymsed case to show that value cycle monitoring can be applied (feasibility), and if applied, what the expected benefits are (usefulness).

Keywords: Data quality · Auditing · Assurance · Information sharing · Supply chain

1 Introduction

Currently, data about shipments available in international supply chains does not provide a timely and accurate description of the goods [1]. For customs authorities, the low quality of data in reporting has proved to be a big problem, e.g. the explosion at the sea vessel MSC Flaminia, where authorities discovered that 605 of the data elements about the cargo in the containers were not accurate. Stakeholders involved have different data formats and communication channels. Also, redundancy and post processing are common problems in measures of control. For example, import declarations and bills of lading about goods are often made several days after the vessels have left the port of origin. Suppose companies in transport logistics and supply chains would be able to improve data quality, then customs can rely on business controls of enterprises, and at least for fiscal matters, additional inspecting and correcting customs related data afterwards at the port of destination would be unnecessary. Data quality issues often result from other stakeholders, further upstream in the supply chain. Under these circumstances, the so called 'push-left principle' [2] could be a solution: the consequences of deviations that are found in an audit or inspection, are 'pushed left', i.e. upstream in the supply chain to the party that caused them.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 467–478, 2015.
DOI: 10.1007/978-3-319-25013-7_38

How can we identify and develop new value adding services and accounting information systems design principles for enterprise, legislative and the audit profession community to solve these data quality problems and achieve sustainable collaboration in international supply chains? This involves enhanced cooperation between different stakeholders. Their bonds and connections are enforced by contracts. However, the principal-agent problem cannot be ignored [3]. Moral hazard and adverse selection[1] problems are sometimes inevitable. Therefore, assurance over data quality is needed, in particular for accounting information systems that are used to record such data [4].

Data Quality Management (DQM) entails the establishment and deployment of roles, responsibilities, policies, and procedures concerning the acquisition, maintenance, dissemination, and disposition of data [5]. We identify some specific challenges concerning DQM in international supply chains and provide some solution guidelines based on a case scenario afterwards in this paper.

Firstly, roles and responsibilities of different stakeholders involved needs to be analyzed, from the manufacturer, exporter and forwarder to the warehouse keeper, customs agent, cargo packers, etc. Secondly, cross-organizational boundaries are difficult to delineate. For example, customs import formalities which are formally the responsibility of the importer, are outsourced to his freight forwarder or customs broker. Thirdly, the contractual relationship is often weak and difficult to manage. For example, the importer depends for his import declaration on data about the goods provided to him by the ocean carrier, but this carrier only has a contractual relationship with the freight forwarder of the exporter of the goods in the country of origin. Business processes and data governance processes need to be well integrated. But contracts are often negotiated on price, not on service level. Other challenges, like the allocation of financial and human resources, require more cost and benefit analysis.

How can we ensure quality of data reported to customs with these challenges? Business reality can generally be modelled as a value cycle: an interrelated system of flows of money and goods [6]. The flow of money should mirror the flow of goods, but in reverse. The point of an accounting information system is to accurately and completely capture these two reverse flows using accounts information. Value-cycle models are well established in the owner-ordered audit tradition in the Netherlands that concentrates on financial reporting completeness, in addition to correctness [7]. When applied to data quality management, value-cycle models can prove to be beneficial and this will be illustrated in this paper.

Our goal here is to propose guidelines for designing and developing an information infrastructure and technology-based mechanism in international supply chains, for data quality monitoring. In this paper, data quality needs are assessed and evaluated within the context of organizational strategies, supply chain structure and existing business processes. First, we provide a brief overview of data quality and relevant definitions, as well as the general steps of data quality assurance. We then apply the value cycle approach to DQM in international supply chains. The next section is an application scenario of a real case in the Netherlands. The paper concludes with recommendations

[1] It refers to a market process in which undesired results occur when buyers and sellers have asymmetric information; the "bad" customers are more likely to apply for the service.

and implications on design principles for implementing value cycle (customs) controls in supply chains.

2 Data Quality Assurance and Data Quality Management

2.1 Defining Data Quality

Data quality is conformance to valid requirements. We should first [5] determine who set the requirements, then determine how the requirements are set. After that, determine the degree of conformance that is needed. In international supply chain domain, both the business and customs need to set data quality requirements. IT organizations/departments need to ensure that the business and customs can have accurate reporting data. They are aware of the existing data quality deficiencies, also the possibility and cost of overcoming them. Sometimes, changes in business processes are needed to address data quality problems. These factors must enter decision process.

Operationally, we can first define data quality in terms of data quality parameters and data quality indicators [8]: A **data quality indicator** is a data dimension that provides objective information about the data. Source, creation time, and collection method are examples. A **data quality parameter** is a qualitative or subjective dimension by which a user evaluates data quality. Source credibility and timeliness are examples. The value is directly or indirectly based on underlying quality indicator values. User-defined functions may be used to map quality indicator values to quality parameter values. For example, if the source is a RFID database, an auditor may conclude that data credibility is high. A **data quality requirement** specifies the indicators required to be documented for the data, so that at query time users can retrieve data within some acceptable range of quality indicator values.

2.2 Dimensions of Data Quality Management (DQM) Objectives

Under general accounting settings, data quality should improve from these dimensions [9]: (a) accuracy/correctness (b) completeness (c) timeliness (d) consistency, etc. For information system and IT infrastructure settings, there are more goals of DQM: (a) integrity (b) independency (c) relevance (d) confidentiality, etc.

2.3 General Processes of Data Quality Assurance (DQA)

Data quality assurance is the process of verifying the reliability of data. Protocols and methods must be employed to ensure that data are properly collected, handled, processed, used, and maintained at all stages of the scientific data lifecycle. This is commonly referred to as 'QA/QC' (Quality Assurance/Quality Control). QA focuses on building-in quality to prevent defects while QC focuses on testing for quality (e.g., detecting defects) [10]. To improve data quality, it is necessary to improve the linkage among the various uses of data throughout the system and across all business process:

1. **Data acquisition and identification**: The first step is to identify critical data areas. Normally this is manifest in two areas: (a) the basic business processes and (b) support for decision making about management of these business processes [11].

2. **Data discovery and profiling**: Data profiling is the systematic analysis of data to gather actionable and measurable information about its quality. Data discovery is achieved by executing data profiling and data monitoring tasks, analyzing data and determining the business rules used to populate the data.
3. **Data cleansing and enrichment**: Detect and correct erroneous data and data inconsistencies both within and across systems. Data enrichment involves enhancing existing data, by adding meta-data or changing data from industry standards and business insights to make it more useful downstream.

3 Data Quality Assurance with the Value Cycle Approach

3.1 Data Quality Assurance (DQA) in Supply Chain Management (SCM)

The relation between DQA and SCM is crucial, stakeholders in supply chains depend on each other, therefore information about agreements and situations must be reliable. Supply chains are generally present in enterprises across logistics, retail and other sectors. In these sectors, supply chains are crucial for business operations and SCM has a significant effect on business performance. "Supply Chain Management describes the discipline of optimizing the delivery of goods, services and related information from supplier to customer" [12]. *Enterprise Resource Planning* (ERP) systems are seen as the digital backbone for information in supply chains, especially when the supply chains are integrated over several companies or departments [13]. There are other information systems as well. Therefore, data monitoring and quality control in SCM can be continuous and automated throughout the whole DQA processes.

In SCM, data quality can have strong effects on operations. Consider for instance the bullwhip effect, which is the phenomenon of amplifying demand variability when moving up the supply chain, leading to growing inefficiencies and diminishing revenues [14]. This means that if a certain piece of data in a supply chain is erroneous or uncertain, fluctuations are increasing rapidly along the supply chain. This effect is affected by data quality, as business processes rely on data provided by others. When data cannot be relied on, it is prudential to keep extra stock. The next link in the chain will think likewise, amplifying the effect. When no specific requirements are set for a certain data element in the supply chain, this will not only cause an overall low data quality, but also amplified variances in stock levels along the supply chain.

3.2 Data Quality in Flows of Money and Goods

Supply chains have a big impact on organizations and are represented by the following flows [12]: goods flow as primary processes, information flow as CRUD (Create, Read, Update, Delete) processes, as well as financial flow. How is information flow linked with goods flow and financial flow? By CRUD operations in a database, whenever the status of the goods or money in the actual flow changes, information changes as well. CRUD processes are the four basic processes that can be performed with data in databases and describe the state of the data at a certain point in the process [15, 16], e.g.

import status changes have to do with the flow of money, as well as the flow of goods through the supply chain. These flows can often be used for cross verification. For establishing proper DQM in supply chains, it is thus required to take into account the goods flow and financial flow. The process of payments depends greatly on data quality, as errors in data can damage the relations with customers.

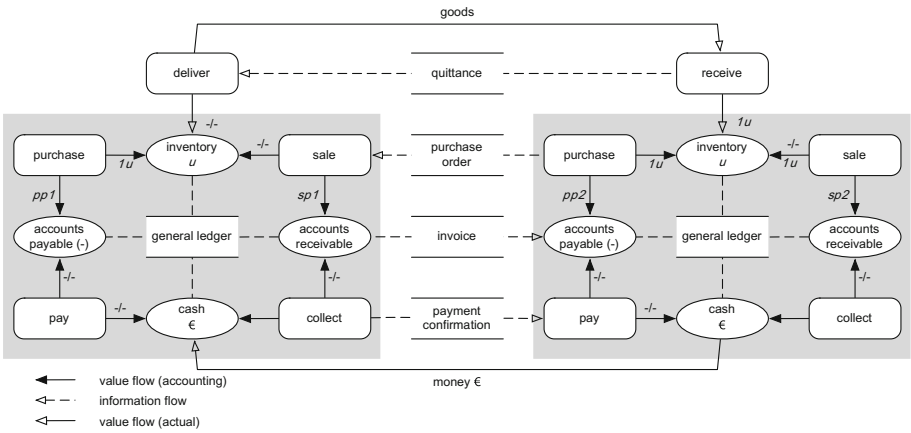


Fig. 1. Value cycle models of two trading companies linked by trade documents [18]

We look at data quality issues from an accounting perspective. Accounting has practices to deal with cross verification, using independent sources of data. One of these practices is to model businesses as a value cycle. Figure 1 shows an example of the value cycle for two trading companies, connected by trade documents (e.g. quittance, invoice, purchase order). We use the following notation. Activities are shown as a rectangle, such as a sales event. Ovals are the recordings of a state of a certain value to the company, such as inventory or accounts payable. States, i.e. accounts, are related through reconciliation relationships, indicated by dashed lines, which come together in the general ledger. The direction of the arrow indicates the influence of events. Arrows generally indicate an increment, while the sign '–/–' indicates a decrement of the corresponding account. Thus, a purchase leads to an increment of the accounts payable, while the purchased goods are added to the inventory. A sale leads to an increment of the accounts receivable and a decrement of the inventory, and so on.

The general idea of value cycle modelling is to use *Reconciliation Relations* to define a mathematically precise model of how the flow of money and goods should be (SOLL), depending on the specific manufacturing inputs and outputs for each type of business, and use it to verify actual audit business samples against (IST) [7]. The mathematical models could be instantiations of the following kinds of equations.

In Fig. 1, for all accounts *S, T* that are affected by an event *e*: $(S) \leftarrow [e] \rightarrow (T)$, e.g. $(inventory) \leftarrow [sale] \rightarrow (accounts\ receivable)$, we have the following transformation equations, where *f* is a constant (here *sales price*) that depends on the business model:

$$input\,(T, e) = f \bullet output\,(S, e)\,,$$

As we record accounts in specific units of measurement (kg, 22 ft container, $, mph), we also need conversion equations:

$$T \text{ in unit } u = f \bullet T \text{ in unit } v \text{ [17].}$$

In addition, for all accounts S, we have the following preservation equation:

$$S(t_1) = S(t_0) + input(S, [t_0, t_1]) - output(S, [t_0, t_1]),$$

where for time interval $[t_0, t_1]$, input(S, $[t_0, t_1]$) $=_{def}$ Sum(input(S, e), for e in $[t_0, t_1]$).

3.3 Applying the Value Cycle Model to Customs Reporting

We need to adjust the value cycle model in three respects for international supply chains. First, add costs components related to goods transport and handling. Second, verify across inter-organizational links. Third, the key approach is finding the right reconciliation relationships that govern the international supply chain domain, in particular, capturing equations related to the flow of physical goods [18].

Here is a specific example of the goods flow in a bonded warehouse. Figure 2 illustrates that the data about goods entering a bonded warehouse[2] should correspond, according to many reconciliation checks, with the data about goods leaving the warehouse, either in transit or for import into free circulation in the EU. Customs have delegated controls over the warehouse to the company. To make sure the warehouse management system is reliable and no goods or documents are missing, they verify this afterwards every month, on the basis of electronic data. This is called 'electronic declaration' or 'audit file'. The so-called 'stock movement declaration' is part of the electronic declaration. The basic principle is that the total in the movement of goods must be balanced, using the following preservation equation, for any period of time:

$$opening\ balance(BV) + entries(BI) - debit\ entries(AF) - closing\ balance(EV) = 0.$$

From this formula, we can derive more equations for the case of boned warehouse:

opening stock at the beginning of a calendar year + entry of goods + internal changes = closing stock at the end of same calendar year + removal of goods;
opening stock at the beginning of a calendar year + entry of goods – destruction – vaporization–loss + findings + adjustment + other = closing stock at the end of same calendar year + goods released for free circulation + re-export + other.

What makes the above equations complicated is the case when several information lines together make up one mutation. For example, three different articles (A, B, C) are packed together into one article (D) (three-in-1 box) according to the rules governing usual forms of handling. In the stock records, the individual articles are registered separately from

[2] The bonded warehouse is under responsibility of a company, and used to store their goods under customs supervision, requiring a formal license from customs to operate. Until a customs destination is known, e.g. re-export (transit) or import (free circulation) no import duties are due. (See also https://en.wikipedia.org/wiki/Bonded_warehouse).

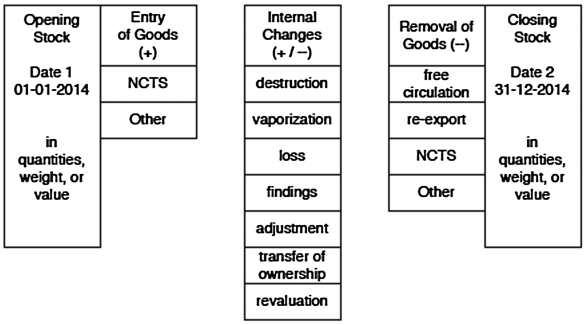


Fig. 2. Conceptual model of the inflow and outflow of a customs warehouse

the articles packed together (different article codes). When the three articles are packed together, this has an effect on the individual stock levels. The stock of the different articles (A, B, C) reduces and the stock of the article packed together (D) increases. This therefore involves 4 transactions: 3 minus-transactions and 1 plus-transaction. These information lines are linked to each other by including the same mutation number of each information line in the declaration system.

4 Application Scenario

4.1 Data Collection

Data for this application scenario description was collected from relatively open interviews with experts in trade, customs legislation, customs audit and companies. The materials from these interviews are supplemented with public sources about customs legislation and inspection policies.

4.2 Scenario Description

ABC is a Dutch company manufacturing machines for international market. Many components in manufacturing are produced abroad. Some of these parts are built into appliances and re-exported within or outside the EU. Other parts are re-exported directly. In the last case, no customs import duties are required. For these reasons, some warehouses at ABC operate as a Bonded Warehouse. Under this license it is allowed to store components from foreign origin, of which payment of import or export duties has been suspended, for an indefinite period until a customs destination is known: entering into the EU (free circulation), or re-exporting outside the EU (transit). In order to obtain and keep the license, ABC must allow regular audits and inspections from the Dutch customs. These involve IT audits of the administrative systems of the warehouse, physical security measures and occasionally inspections to find out if the inventory in the bonded warehouse corresponds to the reported goods.

We are interested in the requirements on ABC's inventory management and information systems. The most important requirement is that customs authorities should be

able to rely on ABC's administration. It should be impossible to lose track of the stored goods (accurate and complete recording), to lose goods from the warehouse (physical security), or to assign the wrong customs destination (procedures). Errors that have a negative impact on accuracy or completeness of reporting data directly affect these key objectives. Data quality is crucial in two capacities:

1. Data quality is part of the requirements that follow from the Customs Warehouse license. It is therefore a key characteristic that must be audited regularly. Both the company and the auditors of the customs office therefore develop policies and procedures for assessing reliability of the company's record keeping.
2. In the audit itself, audit evidence is used and produced by the party being audited, so its reliability is not immediately guaranteed. Therefore, also to improve data quality for its own purposes, the company must build in certain additional precautions into the business processes, procedures and information systems to ensure reliability: so-called internal controls, e.g. segregation of duties; master data management; limited reading and writing access rights; access control measures; logging and monitoring; baseline security. Many controls are implemented using IT systems, hence, IT audit is necessary. These controls also need to be audited.

4.3 Issue

As part of the regular audits, both customs and the internal control department of ABC have now identified a number of weaknesses in the internal controls. Initial analysis has revealed that many errors can be attributed to the crucial process of 'reception', when goods are entered into the warehouse. At this point, ABC can still ensure that records of the goods match with the actual contents, on aspects like order numbers, container numbers, price, origin, goods description and bar-codes. Once goods have entered into the warehouse, it is much harder to trace errors and mistakes. For this reason, ABC has temporarily implemented a number of manual checks regarding the reception of goods. These controls are meant to detect and immediately correct deviations between shipping documents, such as the purchase order, invoice, customs declaration, the actual goods being received at the warehouse, etc.

What complicates the issue is that the bonded warehouse, usually consists of many physical warehouses also contains goods that are not under customs supervision.

Another complication is that at some of these premises, the reception process is not carried out by ABC itself, but has been outsourced to logistics service providers. So, ABC is dependent on logistics providers to carry out these checks adequately.

4.4 Solution Analysis

ABC has identified risks of overall processes, from general IT control, setting up purchase order and production order to sales. Also using controls in their ERP system and prescribing the right sequence of procedures in ERP, ABC sets up controls based on risks. The risk matrix ABC identified is updated on a regular basis. Experts from ABC jointly with the customs made the following steps to produce the risk matrix:

1. Identify for each individual movement type in ABC's ERP whether it is customs relevant or not. Not customs related means end of data flow, so if the goods are mistakenly categorized, it should trigger a control response in the system.
2. ABC implements for each identified customs relevant movement type a specific internal control to mitigate that risk.
3. These internal controls are built in ABC's ERP system. Financial flows rely on logistics flows and are fully automated in ERP.
4. ABC arranges regular IT audits to ensure these controls are working well in ERP.

The monthly declarations of ABC for customs are made using *Automated Periodic Reporting* (GPA), which is generated by a special information system. The EU has a special system *New Customs Transit System* (NCTS) for the reporting of so-called transit goods; i.e. goods that have entered the EU via a specific country, but have not yet been formally imported, and hence, for which no import duties have been paid yet. NCTS requires a manual step to enter the transit status of goods, and then returns a specific *Movement Reference Number* (MRN), which can be used as proof that these goods have the transit status. MRN is essential for ABC to generate an accurate GPA about goods. This manual reporting is done by the freight forwarder FF that has arranged the transport of their goods into the Netherlands.

We propose a systematic approach to improve internal controls, based on general risk management approaches (e.g. COSO ERM). First, identify remaining deficiencies in data quality; these may indicate risks related to customs compliance. Second, find the underlying root cause of these deficiencies. In many cases, the cause will be with another party, on whose data the organization depends. Try to fix deficiencies by improving information systems, processes or even conditions in the contract with other parties. Third, evaluate the remaining compliance risks. If they are unacceptable then repeat the procedure. This approach is called 'push-left' principle [2], because it aims to push any remaining control deficiencies left in the supply chain.

How can we measure the residual risks in a manageable level and indeed 'push left'? Process control in data processing is the underlying basis for data quality. If there is a gap in that process, this could mean goods are disappearing in this case. Only after being fully in control of the data processing can we go to the next level to see if the data is correct. If data at the next level is wrong, it might have financial impact but this can be fixed afterwards. Process control is on top of data quality problem, and is more about optimizing the physical goods' movement. From GPA to the risk matrix, assume which fields are mapped and covered by standard procedures, and then scope into a customs related risk matrix. After that we can delve into data quality. The suggested process controls for data quality assurance (DQA) are as follows:

1. ***Acquire and record data from various sources with segregation of duties***.

Get data from different sources with adequate controls. Segregation of duties before data collection is a precondition for DQA at the source company. Despite internal controls, third parties who manage the information should also be unbiased.

Value cycle monitoring, as represented in Fig. 1, can play a crucial role in the analysis of the segregation of duties. The key auditing question from a customs point of view is

how ABC can assure the accuracy of reporting data. This is an issue, because of the chain dependency of ABC on FF in providing relevant data. Also due to the manual processing of the transit status of goods transported by FF to the bonded warehouse of ABC, mistakes can happen. We will now explain how the model in Fig. 1 can be applied to analyze this auditing problem.

First, these transit status reports of FF can be viewed as an information service provided by FF to ABC. Actually this information service is just one activity in a broader portfolio of information services called customs brokerage, which are typically provided by FF and customs brokers.

The second observation is the chain perspective. The key assumption of the model is that data accuracy can be improved by using the countervailing interests between the different parties in a value network. In this case the value network consists of a simple chain of two parties: FF and ABC. FF has a different interest than ABC, because, although FF offers the transit status report as a commercial service to ABC, it does not directly affect FF's own business interest if they made a mistake. But the accuracy of these data is of direct interest for ABC, because they need to be compliant to customs, and if the report of ABC were not correct, there would be risks of being fined by customs. Therefore ABC added extra controls in ERP to double check whether the transit status reports that they receive from FF are accurate, accuracy of these data improve the accuracy of ABC's reporting to customs.

Thirdly, from a customs auditing point of view this chain can be viewed as a typical example of segregation of duties, as is depicted in Fig. 1, which enhances data accuracy. Another party, namely ABC is double-checking FF who is producing the transit status reports; the whole chain receives a positive audit assessment, because of the built-in segregation of duties between FF and ABC for this data validation.

2. ***Validate data at the source against predefined data quality requirements.***

Evaluate those manual checks, set more explicit data requirements in the contract with the vendor. The 'Push left principle' requires more responsibility from the vendor.

Develop automated services for validating data records at the source. A strategic implementation enables the rules and validation mechanisms to be shared across applications and deployed at various organizations' information flow for continuous data inspection. These processes usually result in a variety of reporting schemes, e.g. flagging, documenting and subsequent checking of suspect records. Validation checks may also involve checking for compliance against applicable standards, rules, and conventions. A key stage in data validation and cleaning is to identify the root causes of the errors detected and to focus on preventing those errors from re-occurring [18].

3. ***Set up unified standards, data formats and communication channels.***

All data providers need to agree on a communications protocol and the data format, to standardize data. For example, automated checks are performed during the sending of the GPA to customs. The format of data required for filing is a unified standard, and should be the same tracing back to the source manufacturers. The consequence of lack of IT and data interoperability across all stakeholders in a supply chain is that the process halts and the declarant is not informed.

4. ***Build an information infrastructure to share data between stakeholders. Create a data pipeline with built-in controls, allowing more real-time collaborations.***

After negotiating with different parties involved, dedicate IT resources to build the information infrastructure and share data between all parties in the international supply chain. Transport conventions, systems, procedures and data in the Logistics Layer dominate the management of the supply chain. But the data relating to the goods to be bought, sold and moved needs to be known in the Transaction Layer to ensure the order is properly met and paid for. If that information was clarified and verified at the point of consignment completion and captured in a data system running parallel to the Logistics Layer then many of the risks associated with poor data would be reduced [11]. This means for reports about goods entry into the bonded warehouse, collect data via the data pipeline from the actual packing list of the consolidator that actually 'packed the box' with goods in the country of origin.

For automated monitoring and sufficient build-in controls, an application platform should include much more than a traditional server operating system does, e.g. a modern cloud platform could provide capabilities such as data synchronization, identity and entitlement management, and process orchestration[3]. The platform should also provide access to new technologies and ideas of enterprise computing.

5. ***Check reconciliation relationships and build feedback systems to better monitor.***

With the help of normative or prescriptive equations in Sect. 3.2, deviations in the actual flows of money and goods can be identified based on actual measurements of the variables during operations. The checks could be on the net weight, number of units and money value using the equations we illustrated in Sect. 3.3. Re-valuation and transfer of ownership also need to be carefully checked with details.

Meanwhile, create automated feedback loop with human capital investment. If one data user (either internal or external) detects a data defect, he can create a flag in the system and the defect will be automated sent to the source for reviewing.

5 Conclusions

How can we get quality data with multiple standards, formats and communication channels in international supply chains? How can the value cycle approach contribute to data quality management for customs reporting? To what extent can we reduce the redundant manual checks and costs of control in data quality management?

In this paper we tried to answer these questions by introducing an approach that builds on value chain modelling from a chain perspective to application in international supply chains. This approach, specifically for data quality assurance in customs reporting, is based on segregation of duties and developing verification equations that can be used to verify data quality across the whole supply chain.

[3] See more on www.thesupplychaincloud.com and www.opengroup.org: Cloud Computing Open Standards, *the Supply Chain Cloud Report.*

We illustrate the approach by a case scenario of a manufacturing company. It shows that the steps of data quality assurance we proposed can be implemented and if successfully would be beneficial for different stakeholders. If the goods information generated by the vendor, at the starting point of the supply chain, is accurate and complete, those manual checks by parties at the other end of the supply chain would be unnecessary. This is only part of the research, and we leave the development of analytical detection models from reconciliation relationships for further research. Nevertheless, we believe that if information sharing could be improved this way, data quality in international supply chains could also be improved and regulatory compliance risks would be reduced, resulting in operational benefit as well.

References

1. Hesketh, D.: Weaknesses in the supply chain: who packed the box? World Customs J. **4**(2), 3–20 (2010)
2. de Swart, J., Wille, J., Majoor, B.: Het 'push left'-principe als motor van data analytics in de accountantscontrole. Maandblad voor Accountancy en Bedrijfseconomie **87**, 425–432 (2013)
3. Eisenhardt, K.M.: Agency theory: an assessment and review. Acad. Manage. Rev. **14**(1), 57–74 (1989)
4. Romney, M.B., Steinbart, P.J.: Accounting Information Systems, 10e. Prentice Hall, Upper Saddle River (2006)
5. Geiger, J.G. (ed.): Data quality management the most critical initiative you can implement. In: The Twenty-Ninth Annual SAS® Users Group International Conference (2004)
6. Starreveld, R.W., de Mare, B., Joels, E.: Bestuurlijke Informatieverzorging (in Dutch): Samsom, Alphen aan den Rijn (1994)
7. Blokdijk, J.H., Drieënhuizen, F., Wallage, P.H.: Reflections on Auditing Theory, a Contribution from the Netherlands. Limperg Instituut, Amsterdam (1995)
8. Wang, R.Y., Kon, H.B., Madnick, S.E.: Data quality requirements analysis and modeling. I: Ninth International Conference of Data Engineering; Vienna, Austria (1993)
9. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manage. Inf. Syst. **12**(4), 5–33 (1996)
10. Chapman, A.D.: Principles of Data Quality. Global Biodiversity Information Facility, Copenhagen (2005)
11. Orr, K.: Data quality and systems theory. Commun. ACM **41**(2), 66–71 (1998)
12. Cooper, M.C., Lambert, D.M., Pagh, J.D.: Supply chain management: more than a new name for logistics. Int. J. Logistics Manage. **8**(1), 1–14 (1997)
13. Gunasekaran, A., Ngai, E.W.: Information systems in supply chain integration and management. Eur. J. Oper. Res. **159**(2), 269–295 (2004)
14. Lee, H.L., Padmanabhan, V., Whang, S.: The bullwhip effect in supply chains. Sloan Manage. Rev. **38**(3), 93–102 (1997)
15. Martin, J.: Managing the Data Base Environment. Prentice Hall, Upper Saddle River (1983)
16. Polo, M., Piattini, M., Ruiz, F.: Reflective persistence (Reflective CRUD: reflective create, read, update and delete). In: Sixth European Conference on Pattern Languages of Programs (EuroPLOP) (2001)
17. Veenstra, A.W., Hulstijn, J., Christiaanse, R., Tan, Y.-H.: Information exchange in global logistics chains: an application for model-based auditing. In: PICARD2013 (2013)
18. Redman, T.C.: Data Quality: The Field Guide. Digital Press, Boston (2001)

Towards a Federated Infrastructure for the Global Data Pipeline

Wout Hofman(✉)

TNO, The Hague, The Netherlands
Wout.hofman@tno.nl

Abstract. Interoperability in logistics is a prerequisite for realizing data pipelines and the Physical Internet. Forecasting data, real time data, and actual positions of shipments, containers, and transport means shared via events have to be harmonized and are expected to improve all types of processes, support synchromodal planning, and improve risk analysis from a compliance and resilience perspective. Technically, several solutions are implemented by organizations and innovations have been validated in so-called Living Labs or demonstrators in various projects. These solutions do not yet provide open systems required for a (global) data pipeline. A federation of solutions is required to construct data pipelines and to support sustainable development of applications on smart devices allowing Small and Medium sized Enterprises to collaborate. This paper proposes a set of platform services and so-called platform protocols to allow interoperability of different platforms for constructing a data pipeline. The proposed services and protocols further extend existing interoperability solutions and services for supply and logistics.

Keywords: Seamless interoperability · Data pipeline · Federated platforms · Service · Protocol · Physical internet

1 Introduction

Customs authorities require additionally data to their current declaration for risk analysis improvement and introduced the concept of data pipeline for seamless data sharing as a solution [1]. Such a data pipeline consists of a large number of stakeholders like shippers, consignees, forwarders, and carriers, exchanging value according a transaction hierarchy, called logistic chain, in an organizational network (reference). The actual implementation of a data pipeline is by interconnecting legacy systems of the stakeholders and/or support by commercial – and community solutions [2]. It is not to be expected that one global system will implement the data pipeline, but interoperability between existing systems and solutions needs to be constructed [2]. As of currently, many interoperability implementations in trade and logistics are based on the message paradigm, but also other mechanisms are explored to address for instance real time data sharing for dynamic planning or resilience [3], Service Oriented Architecture (SOA) supported by Enterprise Services Busses (ESBs) [4] or Linked Data [5]. For real time data sharing, the current generation of platforms supports an Application Programming

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 479–490, 2015.
DOI: 10.1007/978-3-319-25013-7_39

Interface (API) registry [6], a particular SOA implementation based on the REST protocol. APIs are still technical specifications that require interpretation to derive semantics.

Seamless data sharing between systems and components of different stakeholders requires universal connectivity [7]. In this respect, scoping of specifications is also important like taking a bilateral or multilateral interoperability [8] or a modeling approach covering organizational chains [9]. Interconnecting internal business processes resulted in reference models either specifying both processes and data [10] or only data with a messaging choreography [11]. Implementation of these reference models still lead to closed systems, since, organizations make bilateral or community agreements based on these models [12]. Several sources [13, 14] stress the importance of unambiguous semantics as part of interoperability, but do not address the implementation of this semantics in legacy systems or other solutions. It is yet unclear how process aspects need to get addressed in interoperability. Interoperability layering [14] considers pragmatics without presenting a way to model pragmatics like taking the bilateral or multilateral, chain approach.

A complicating factor is that Small and Medium sized Enterprises (SMEs) cover some 80 % of the logistics market [15] performing some 20 % of the business. These SMEs have either simple or no IT solutions or systems, but interface manually with systems of their customers, potentially supported by web interfaces. Thus SMEs have to deal with different interfaces to become interoperable with their customers instead of having simple applications running on smart devices with cloud solutions of one or more communities and or providers, since these SMEs operate international and require interfacing with many systems and solutions.

This paper proposes a set of platform services that enables an enterprise to connect once to an infrastructure of federated platforms and compose a data pipeline. Standardization of this set of services allows development of applications on smart devices, where these applications can interconnect to any given platform thus creating a sustainable business model for app developers in logistics. Each solution provider in this infrastructure can have its particular implementation of the services, thus satisfying their customer requirements and have sufficient market share. Firstly, requirements to platform services leading to design choices are formulated and secondly the services and the protocol for platform federation are introduced. The research presented by this paper is based on an action design research approach [16] across several EU funded and Dutch projects addressing interoperability in logistics. Each project has constructed artifacts that do however not meet requirements formulated in [7, 17].

2 Design Choices

A (federation of) platform(s) has to meet particular user requirements. Since it is fairly complex to assess user requirements for all global data pipelines, those stemming from various European Union (EU) funded projects and literature will be transformed into design choices. An example of a design choice is for instance the support of the messaging paradigm, common to most interoperability implementations between organizations. By making

these design choices explicit, discussion on their applicability to meet user requirements is supported. Design choices are on distinction of 'service', 'protocol', and 'interface', bilateral versus multilateral business process modeling, semantics, and data governances supported by privacy-enhanced technologies. This section presents choices based on practice inspired research [16] and briefly reflects the state of the art in research. It does not pretend to be complete, but identifies some basic research questions. The answers to these questions have to be supported by a federation of platforms; the next section shows the mechanisms to do so. Semantics is core to all choices made.

2.1 Participating in a Federated Infrastructure

Currently, organizations bilateral or multilateral develop interoperability agreements, encompassing both functional and non-functional aspects, like message implementation guidelines and process alignment, based on open standards or with proprietary formats leading to closed solutions [12]. Each time a business relation with another enterprise needs to be established, investments in agreements has to be done. Seamless interoperability [7, 17] addresses this problem, but does not provide solutions. [9, 18–19] introduce business process modeling either for bilateral or multilateral interoperability as a solution, but [20] argues to model only behavior between any two peer entities. There are different ways to specify behavior; [21] provides transaction templates for bilateral interoperability to construct chains. A generic specification of behavior will not be applicable to all resources, since they all have different goals and capabilities [22]. A generic specification can however serve as a reference framework for specifying these particular goals and capabilities.

This paper proposes to apply the concept of 'resource' offering both real time data and providing or requesting logistic services as specified by an 'Information Profile' of such a resource. In this respect, several issues need to be addressed, namely how to express the external behavior of a resource in terms of interactions and semantics. The concept of transaction templates to express external behavior for business transactions can be applied [21]; other mechanisms like events might be required to share any logistic state changes like arrival of a vessel and delivery of a container at its destination. Semantics of one's profile can be expressed as an ontology, based on a networked ontology of logistics concepts and services (see for instance ontology.tno.nl for a logistics ontology). Semantics of data and behavior need to have a technical binding to a paradigm like messaging and SOA [4] supported by a syntax like XML or JSON. In case any two communicating organizations have different technical bindings, binding negotiation and a data transformation function have to be implemented, either by a data provider, a consumer, or as service of the federation of platforms (see next section). In case Information Profiles of any two communicating organizations are based on an identical - or matched semantic models, an on-the-fly technical binding by a platform can be constructed, as long as the federation of all platforms support that technical binding.

The concept 'resource' with its 'Information Profile' needs further research and examples stemming from real use cases. These examples are currently developed in EU funded projects like EU FP7 CORE.

2.2 Data Governance and Privacy Enhanced Technologies

Organizations are hesitant in sharing information due to for instance commercial or liability reasons, e.g. the amount of free capacity of a barge might reduce prices, the location of a truck might increase vulnerability for cargo theft or providing real-time and predicted depth of a waterway might increase liability. There are currently a number of barriers that block the adoption of data sharing amongst resources, e.g. data ownership, privacy, commercial sensitivity, liability, and culture [23]. In this respect, data and events are classified as:

- Open data. Data is publicly available to everyone. Open data is normally considered to be available without any costs, but in some occasions like the Cadastre data, one needs to pay.
- Community data. Data is shared within a community according agreed rules. Like with open data, one might distinguish free - and paid data.
- Partner data. Data is shared with a specific partner.
- Internal data. Data is only shared within an organization, according internal data policies (Fig. 1).

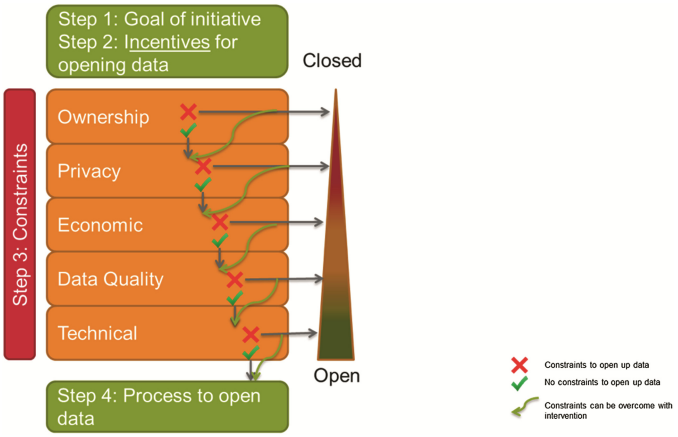


Fig. 1. Decision support instrument for data sharing [23]

Data classification has a lifetime, e.g. it may change over time and/or may be applicable for one or more calls or interactions. For instance, available capacity on a trip may be shared only once in a community at the start of the trip or can be updated during the trip. Communities can also be flexible, e.g. organizations can join and leave a community over time.

The previous figure shows the decision model developed by [23]. It addresses various aspects like data ownership, privacy and commercial sensitivity, and economic aspects,

resulting in a data policy supported by interventions. Many of these interventions are supported by privacy-enhanced technologies [24] like identification and authentication, access control, filtering, and homomorphic encryption [25]. One particular technology might support data sharing for one or more decisions, but we have not yet found these for logistics data sharing. For instance, Role Based Access Control expresses access control for internal data, but a more fine grained access control mechanism like Attribute Based Access Control might be required for partner – and community data. Templates for particular roles can be specified to ease the specification of data policies, e.g. a template for a role like a forwarder. These templates are a form of Role Based Access Control that can thus be refined by organizations meeting their particular requirements as they operate in one or more roles. These templates may implement formal restrictions from a liability and financial perspective, e.g. carriers should not receive any data on the content of a container, whilst they are otherwise liable for any damage or loss to the content.

Role – and Attribute Based Access Control can be expressed as an ontology of a set of rules, based on the earlier mentioned networked ontology for logistics. To support global logistics and supply, protocols are required for a federation of identities [26]. Further research is required with respect to the relation between privacy-enhanced technologies as interventions for data governance and implementing these technologies in real world cases.

2.3 Service, Protocol, and Interface

The Internet design principles of 'service', 'protocol', and 'interface' [20] are applied for specifying a federation of platforms supporting data sharing between organizations. A federated platform is said to offer a 'service' to back office systems of supply and logistic stakeholders, e.g. the ability to exchange messages, validate the message structure and content, and validate the message sequence, whereas a 'protocol' between any two platforms provides the ability to actually share data with for instance messages. The protocol is the set of agreements for sharing data between any two platforms, independent of a local implementation of the service by each of those platforms to their users, which is called 'interface'. The service is the conceptual representation of this protocol to a one or more back office systems and/or end-user. The same service can be implemented by various interfaces, e.g. a file sharing mechanism or an API can serve as an implementation of a service. In fact, 'interface' is the technical binding of a service offered to back office systems of an organization. The technical binding of an interface can differ from that of an agreed binding of the protocol, which requires transformation by a local implementation of the protocol and service by a platform (Fig. 2).

Introducing these concepts allows conceptually specification of a service and a protocol with different technical bindings, both for the service by its interface and a protocol for interoperability between two local implementations. Complexity reduction of federated platforms is achieved if all participating platforms support the same semantics and technical bindings of the protocol.

Each service – and protocol primitive has a particular structure with control information and a payload, where a semantic model specifies the semantics of the payload.

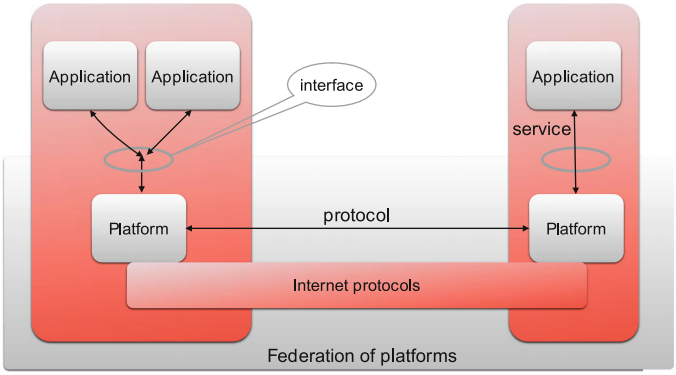


Fig. 2. The concepts service, protocol and interface

Control information of a service primitive is applied by a local implementation for processing that primitive and constructing a protocol primitive with a negotiated - or on-the-fly technical binding (see before). It allows for instance the transformation between a SOA interface of a service to a message based protocol and vice versa. Communication is already implemented by various protocols and is therefore considered out of scope. Thus, the service and protocol for logistics will be elaborated, whereas a local interface is discussed in the next section considering back office integration.

3 A Federation of Platforms

The previous section has introduced a number of research questions and concepts for federated platforms. This section further elaborates the services of a (federation of) platform(s) and its protocol. The services reflect the concept of Information Profile and the support of data governance, all based on networked ontologies.

3.1 Federated Platform Services

A (federation of) platform(s) provides generic services to resources, where these generic services are configured by semantics for a particular application area like supply and logistics. The services can be categorized in two main groups that can be further decomposed (Fig. 3): operational services consider actual sharing of logistics data utilizing various technical bindings and mechanisms; administrative services consider registration and other types of support services for operating a (federation of) platform(s). The decomposed services can be defined as:

- Registration Services. These services support an organization in its registration and connection to a platform (profile specification services) and specify its data policies (data policy specification services).

- Real time data sharing services. These support data policy negotiation, both for events and data, search for a particular (composite) service matching a goal (matching services), and sharing the state of supply and logistics chains via events (visibility services). Visibility services can be on particular objects like trucks with their location, speed, etc. in an area (geo-fencing), timeframe (time-fencing) or a combination of both [27]. Visibility services can only provide the state of an object or also evaluate the state against requirements, which is supported by (complex) event processing [28]. Matching services can be on search for a (structured set of) profile(s) meeting a customer goal, where a structured set composes a logistics chain or all retrieved profiles exactly meet a customer goal. These matching services can be applied in various ways, e.g. to support synchromodal booking [28].
- Transaction support service: validating the sequencing of interaction according an agreed transaction protocol specified by a choreography. In this particular case, the initiation and processing of transactions is in applications registered at a platform.
- Data sharing service: reliable and secure exchange of data and events according a particular technical binding, potentially supported by data transformation. Reliability services consider data resubmission in case the receiving platform or application responds and are not always required. The same is applicable for secure data sharing.
- Supporting services. These services support the operation of a federation of platforms providing particular services to platform users. These services not necessarily have a supporting protocol. The following supporting services are foreseen:
  - Semantic services providing the networked ontologies (see Sect. 2).
  - Publish/subscribe services providing the ability to subscribe to particular events, where these events provide state information. Publish/subscribe may require policy negotiation services.
  - Non-repudiation services that provide proof of actual data shared between any two users of a federation of platforms. Non-repudiation services are supported by an audit trail registering all actions in terms of data sharing between any two actors, e.g. sending or receiving particular data with a timestamp, a log containing the actual data that has been shared, and monitoring services providing both access to the audit trail and the log.
  - Accounting and billing services supporting paid data according agreed pricing structures. These services utilize the non-repudiation services.
  - Certification services providing identification and authentication of platform users.

Services are interrelated. Visibility services can for instance use publish/subscribe services and data exchange services for events, transaction support services can utilize data exchange services by messaging and reliability services to assure a reply is received in time. Secure and reliable data exchange is for instance specified by Electronic Business XML with ebMS [19] and implemented by an eFreight access point [29]. A formal service specification considers the control information and payload [20], which yet needs to be performed.

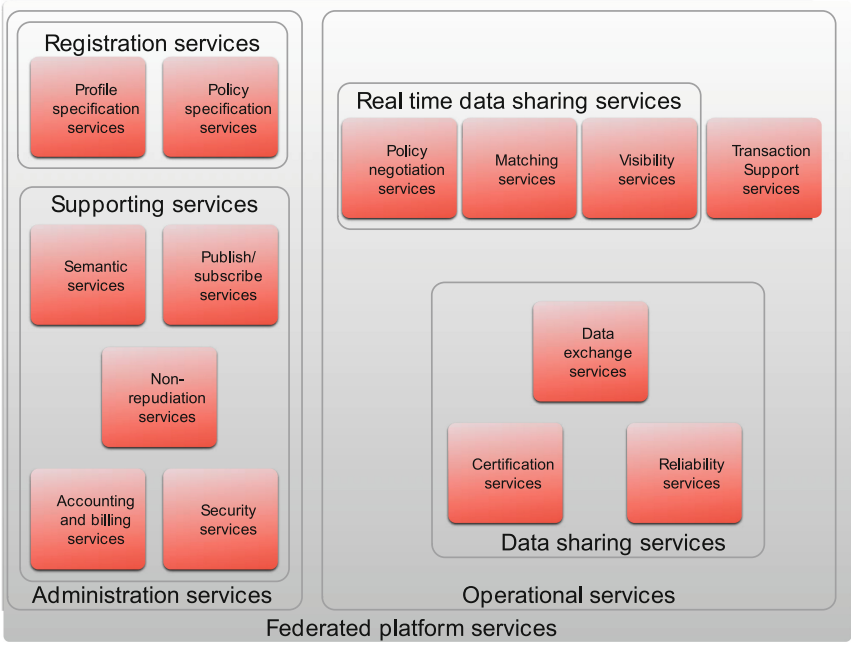


Fig. 3. Federated platform services

3.2 Federation of Platforms

Like the services, the protocol can also be decomposed. The protocol should support each of the services, but some protocols can support more than one service by a different payload of the protocol primitives [20]. The protocol is decomposed as follows:

- Data policy negotiation protocol: negotiate the data that can be shared amongst two organizations.
- Matching protocol: sharing goals and (a structured set of) service(s) to support matching services.
- Visibility protocol: how to access particular data on the supply chain status, e.g. it basically consists of events either received upon subscription or by a query.
- Transaction protocol: the agreed choreography of interactions for transaction support services.
- Data sharing protocol: sharing data between two communicating systems, where the payload is provided by the aforementioned protocols. The data sharing protocol is decomposed in:
  - Binding negotiation protocol: to establish the technical binding for sharing data.
  - Data exchange protocol: the actual sharing of data (and events) according an agreed or selected (on-the-fly) technical binding.

  - Reliability protocol: resubmission and identification of resubmitted messages, potentially resulting in a receipt acknowledgement, and timers to detect timely replies to support reliability services.
  - Security protocol: selecting a secure protocol with agreed certificates, e.g. https.
- Authentication protocol: a protocol for federation between certification authorities to support authentication of an identity [26].

One might consider to introduce a protocol for supporting registration services, e.g. to share complete Information Profiles and data policies. Such a protocol would provide complete transparency. In our proposed approach, the data negotiation -, matching -, and binding negotiation protocol support this type of transparency, but does not provide a generic data policy of any registered user.

3.3 Implementation and Deployment

There are several ways to implement a federation of platforms, e.g. if any two enterprises are connected to the same platform like a Port Community System, data sharing services between them might be implemented via a database and registration is based on administrative services with a proprietary format. Only in case two enterprises connected to different platforms, the protocols will be required. Currently, most of these protocols are message based (see before).

There are two dimensions to the deployment of the Connectivity Infrastructure, namely an business dimension and ICT dimension. Both will be discussed here. The ICT dimension addresses the development of a local implementation, with potentially different components providing different services and supporting particular parts of the protocol. The following options are feasible:

- Open source: the services are provided by an open source software solution that every resource can implement. Like indicated before, eFreight Access Points provide particular functionality, so do iCargo Access Points in supporting virtualization of logistics actuator objects like containers and trucks [3].
- COTS (Commercial Of The Shelve): the implementation of the protocol with a local interface to its back office systems based on COTS. The software offering the service and supporting the protocol is licensed to a resource or its owner that implements the functionality. The COTS provider is responsible for correct (and complete) implementation of the protocol.
- Proprietary: the IT department of an enterprise develops the implementation of the protocol or its implementation is outsourced to an external software developer. The implementation can be based on open source solutions and/or components like available integration brokers. The solution is owned by the enterprise (with the exception of COTS components used for implementation) and the enterprise is responsible for correctness of the protocol implementation.

From a business perspective, the service can be implemented in many ways, for instance:

- Resource Deployment: each resource implements and operates the protocol by itself. The services are internal to the resource, but utilizing an ICT solution implementing the services reduces development costs. The iCargo Access Points try to provide this functionality [3].
- Community Systems: two or more organizations might decide to implement the services themselves with different local interfaces to back office systems and/or end-user functionality. These organizations thus own the community system. Port Community Systems are examples of these types of systems.
- Cloud Platform: the service and the protocol are provided by a platform of a commercial provider. The latter ones can construct their particular solutions on top of the services. Note that this type of solution is identical to telco or other providers offering services with COTS and/or open source solutions for communication protocols.

The services can also be used for open innovation, implying the development of apps for SMEs (see also www.logicon-project.eu). To support open innovation, the services have to be more tailored to a specific target group of end-users, e.g. barge operators are expected to have other services than truck drivers.

A number of services is currently proprietary to a particular solution, e.g. registration of resources at a community system or cloud platform. These deployment solutions do not yet support the protocol, which makes it difficult to find for instance resources in an infrastructure. Furthermore, these deployment solutions have their specific data policies implemented by message implementation guides, which restricts data sharing across these platforms and require all types of transformations between them.

4 Conclusion and Further Research

The paper presents a set of services for federation of platform solutions and services supported by a protocol for logistics and supply. By standardizing supply and logistics services and their underlying protocol, the so-called data pipeline for interoperability in trade facilitation will be enabled. The services and protocol enable each individual object or actor to act as a information resource with particular capabilities for data sharing. By examples, we have mapped functionality to existing components like developed for eFreight and iCargo. Dedicated solutions and services can be developed addressing particular interoperability aspects like virtualization of actuators representing physical objects like trucks and containers, thus also contributing to the concept of the Physical Internet [7]. Standardisation of services of federated platforms also contributes to development of sustainable business models for deployment of apps on smart devices for SMEs. By separating service and protocol, each resource implementing a protocol stack known to the infrastructure will be able to participate, without additional costs and effort for development of bilateral or community guidelines. Each ICT solution – and service provider will also be able to tailor its services to optimally integrate resources in the infrastructure and provide added value like complex event processing and transformations to these resources.

The solutions provided by this paper need to be developed further, including construction of low cost connection to the federation of platforms. We have indicated that lots of existing systems and cloud solutions have a role in implementing the services. Federation of platforms requires additional research into the business models of these platforms, like a sustainable business model of community systems. Not yet, all identified services are fully supported by software solutions.

Acknowledgements. The content of this paper is based on research and development in EU FP7 SEC Cassandra, EU FP7 SEC CORE, EU FP7 INFSO iCargo, EU FP7 INFSO Logicon, several other projects for data sharing in logistics and supply chains, and discussions in various fora like ALICE on the subject of seamless interoperability.

References

1. Klievink, B., Van Stijn, E., Hesketh, D., Aldewereld, H., Overbeek, S., Heijmann, F., Tan, Y.-H.: Enhancing visibility in international supply chains: the data pipeline concept. Int. J. Electron. Gov. Res. **8**(4), 14–33 (2012)
2. Hofman, W., Bastiaansen, H.: A global IT infrastructure improving container security by data completion. In: ECITL, Zaragoza, Spain (2013)
3. Dobler, M., Schumacher, J.: A pan-european ecosystem for intelligent cargo and its impact on emission reduction. In: Transport Research Arena 2014, Paris (2014)
4. Erl, T.: Service Oriented Architecture – Concepts, Technology and Design. Prentice-Hall, Upper Saddle River (2005)
5. Heath, T., Bizer, C.: Linked data - evolving the web into a global data space. In: Synthesis lectures on the Semantic Web: Theory and Technology, Morgan & Claypool Publishers (2011)
6. Hofman, W., Rajagopal, M.: A technical framework for data sharing. J. Theoreticla Appl. Electron. Commer. **9**(3), 44–57 (2014)
7. Montreuil, B., Meller, R.D., Ballot, E.: Physical internet foundations. In: Borangiu, T., Thomas, A., Trentesaux, D. (eds.) Service Orientation in Holonic and Multi Agent Manufacturing Robots, pp. 151–166. Springer, Heidelberg (2013)
8. Schonberger, A., Wilms, C., Wirtz, G.: A requirements analysis of business-to-business integration. In: Fakultat Wirschaftsinformatik und angewandte Informatik Otto-Friedrich-Universitat, Bamberg (2009)
9. Huemer, C., Liegl, P., Motal, T., Schuster, R., Zapletal, M.: The development process of the UN/CEFACT modeling methodology. In: 10th International Conference on Electronic Commerce (IEC), Innsbruck, Austria (2008)
10. Pedersen, J.T.: One common framework for information and communication systems in transport and logistics: facilitating interoperability. In: Golinska, P., Hajdul, M. (eds.) Sustainable Transport. EcoProduction. Environmental Issues in Logistics and Manufacturing, pp. 165–196. Springer, Heidelberg (2012)
11. World Customs Organization. WCO Data model - cross border transactions on the fast track. World Customs Organization (2010)
12. Hofman, W.: Applying semantic web technology to interoperability in freight logistics. In: e-Freight 2011, Munich (2011)
13. European Commission. European Interoperability Framework for European public services. European Commission, Brussels (2010)

14. Wang, W., Tolk, A., Wang, W.: The levels of conceptual interoperability model: applying systems engineering principles to M&S. In: Spring Simulation Multiconference (2009)
15. Transport and Logistics Netherlands (TLN). Transport in figures, TLN (2015)
16. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS Q. **35**(1), 37–56 (2011)
17. Chituc, C.-M., Azevedo, A., Toscano, C.: A framework proposal for seamless interoperability in a collaborative networked environment. Comput. Ind. **60**, 317–338 (2009)
18. Van den Heuvel, W.-J., Papazoglou, M.P.: Towards business transaction management in smart networks. IEEE Comput. Soc. (2010)
19. Kotok, A., Webber, D.: ebXML - the new global standard for doing business over the internet, New Riders (2002)
20. Tanenbaum, A.S.: Computer Networks, 3rd edn. Prentice Hall, Upper Saddle River (1996)
21. Dietz, J.: Enterprise Ontology, Theory and Methodology. Springer, Berlin (2006)
22. Spohrer, J.K.S.: Service science, management, engineering, and design (SSMED) - an emerging discipline - outline and references. In: International Journal on Information Systems in the Service Sector, May 2009
23. Eckartz, S., Hofman, W., Van Veenstra, A.F.: A decision model for data sharing. In: eGov2014, Dublin, Ireland (2014)
24. Lagendijk, I.L., Erkin, Z., Barni, M.: Encrypted signal processing for privacy protection. IEEE Signal Process. Mag. **82** (2013)
25. Gentry, C.: A fully homomorphic encryption scheme. Stanford University (2009)
26. Pruksasri, P., van den Berg, J., Hofman, W., Deskapan, S.: Multi-level access control in the data pipeline of the international supply chain system. In: iCETS, Macau (2013)
27. Merrienboer, S., Veenstra, A.W., van den Haak, W., Tavasszy, L.: Using floating truck data to optimise port logistics. In: 10th Intelligent Transport Systems European Congress (ITS2014), Helsinki (2014)
28. Hofman, W.: Control tower architecture for multi- and synchromodal logistics with real time data. In: ILS2014, Breda (2014)
29. Angholt, J., Wackerberg, M., Olovsson, T., Jonnson, E.: A First Security Analysis of a Secure Intermodal Goods Transport System. Chalmers University, Gothenburg (2013)

Key Design Properties for Shipping Information Pipeline

Jensen Thomas[1(✉)] and Yao-Hua Tan[2]

[1] Copenhagen Business School, Howitzvej 60, 2000 Frederiksberg, Denmark
tje.itm@cbs.dk
[2] Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands
y.tan@tudelft.nl

Abstract. This paper reports on the use of key design properties for development of a new approach towards a solution for sharing shipping information in the supply chain for international trade. Information exchange in international supply chain is extremely inefficient, rather uncoordinated, based largely on paper, e-mail, phone and text message, and far too costly. This paper explores the design properties for a shared information infrastructure to exchange information between all parties in the supply chain, commercial parties as well as authorities, which is called a Shipping Information Pipeline. The contribution of the paper is to expand previous research with complementary key design properties. The paper starts with a review of existing literature on previous proposed solutions for increased collaboration in the supply chain for international trade, Inter-Organization Systems and Information Infrastructures. The paper argues why the previous attempts are inadequate to address the issues in the domain of international supply chains. Instead, a different set of key design properties are proposed for the Shipping Information Pipeline. The solution has been developed in collaboration with a network of representatives for major stakeholders in international trade, whom evaluate it positively and are willing to invest, develop and test the prototype of the Shipping Information Pipeline.

Keywords: International trade · Inter-organizational systems · Information infrastructure

1 Introduction

Research regarding international trade estimates that 40 % of the delays in the lead-time of supply chains for international trade in the large ports are caused by administrative burdens imposed by authorities. Typically data inaccuracy is 50 % for the information reported by businesses to authorities; since these data are used for the risk assessment, this is rather critical. The annual world-wide extra costs due to administrative burdens of crossing borders are estimated in the range 100–500 Billion US$[1].

The organizations in international trade are characterized by utilizing a wide range of communication channels including phone, e-mail, SMS and paper based media for

[1] Cassandra Research Project presentation, 2014.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 491–502, 2015.
DOI: 10.1007/978-3-319-25013-7_40

information and documentation related to the shipments. This creates issues for the actors involved including lack of knowledge about status, information and documentation of shipments. Information exchange in international supply chain is considered extremely inefficient, uncoordinated and far too costly [1]. To address these challenges it's in line with previous research e.g. [2] propose to use a shared information infrastructure (II) for communication of the related shipping information whereby the information can be shared among multiple organizations[2]. The II for shipping information is named the *Shipping Information Pipeline* (SIP). It's the long term ambition to design, build, and test (ultimo 2015) plus evaluate a prototype of the SIP and this paper only covers the initial design of a prototype.

International trade plays an important role in the global economy and is a complex eco-system. The domain of international trade involves up to 40 actors/organizations in a single shipment including both private businesses and public organizations in a complex eco-system [3]. Furthermore, a serious complication in this eco-system is that the information exchange is extremely inefficient and rather uncoordinated. Complete different IT systems and message standards are used on different sides of the oceans. This has led to a practice, where information exchange is based primarily on paper and e-mail with many re-typing and copy-paste operations, phones calls and text messages. Consequently, it's relatively costly and error prone for the commercial parties as well as authorities e.g. inspection agencies in supply chain. Finally, the number of independent organizations and the huge variety of different IT systems make changes extremely difficulty even if the changes seem beneficial for the overall eco-system.

Usage of modern IT to improve the situation in international trade has been researched and attempted previously but none of the solutions have become used at an international scale. The design of an II as the SIP is different from solutions proposed in previous research which built on more traditionally centralized information systems; because the SIP has no central database, organization or control. For the previous solutions almost all communication is based on bilateral information exchanges between two organizations; whereas an II as the SIP enables sharing of information simultaneously among multiple organizations or even sharing the information by publishing it public. The SIP is comparable to the Internet and designed to be on top of the Internet and its standards. The SIP is an internet for shipping information.

In order to design the SIP one of the major players in international shipping decided to engage in research regarding the design properties for the SIP. The major player foresee that it's impossible for a single player to become successful with the SIP and search for ways to establish a collaboration with other partners including authorities in the international trade. To engage other partners the designing and testing of prototypes of the SIP are important. The research involves design theory for information infrastructure and follows the design science method to guide the research.

Accordingly the research question guiding this paper is:

What would be the key design properties of a Shipping Information Pipeline?

[2] Organizations can be public or private e.g. a company, a cooperation, an association, an institution, etc.- an entity on its own; an actor belongs to an organization and can be user of IS solutions. A stakeholder will have a stake e.g. in owning the SIP.

The scope of this research is limited to the containerized international trade bound for the European Union by deep sea, but this is not a severe limitation since the majority (70 to 90 %) of imported goods to the European Union is carried by container ships. Accordingly the other means of transport are not considered here. The largest volume to Europe passes through the port of Rotterdam. Accordingly the research has focused on trade lanes governed by Dutch authorities.

The rest of the paper is structured as follows: First theories behind previous proposed solutions, Inter-Organizational Systems [4] and design of Information Infrastructures (II) are presented. Secondly the research method is described. Thirdly are the proposed key design properties for a domain specific II for shipping information within the supply chain for international trade for containerized trade bound for European Union. Finally, an evaluation is provided and followed by a discussion.

2 Literature Review

The roots of the idea to share data via an electronic pipeline for international trade and subsequently attempts to develop IOS for international trade are presented below. Followed by a presentation of theories regarding IOS solutions currently used for information exchange and the Design Theory and design properties of II.

2.1 The Roots of the Idea of an Electronic Pipeline for International Trade

The idea of a sharing data via an electronic pipeline roots back to two official EU representatives from Dutch and United Kingdom customs. They came up with the idea of an electronic pipeline when visiting China, and they proposed that the European authorities should get the data about each and every shipment from the source e.g. the Chinese company packing the container. This kind of solution would dramatically reduce import authorities' serious problems with poor data quality; which is seriously affecting the customs/inspection authorities, when they are performing risk assessment of inbound cargo.. Several studies have identified that the quality of the data provided to authorities is poor, misleading and in some cases even fraudulent. An illustrative example is that when a container vessel stranded, there was a comparison of the filed information on the contents of the containers and the actual contents when the containers were opened and the authorities found that the data quality was only around 60 % [5]. Accordingly, there is a clear advantage to the EU authorities, if they via an electronic data pipeline could get the data directly from the source.

This idea caught on in the EU, and research funds were made available to fund a range of research projects addressing some of the technological challenges and demonstrating the possible solutions. Important projects addressing the issues and analyzing the potential benefits for organizations involved in international trade: ITAIDE[3], Contain[4], Integrity[5],

[3] www.itaide.org.
[4] www.containproject.eu.
[5] www.integrity-supplychain.eu/.

Cassandra[6], and iCargo[7]. In the ITAIDE project (2006–2010) an "I3" framework was developed for accelerated trade through networks of trusted traders utilizing an II built on IT innovations which enables four critical capabilities[8]: Real-time monitoring, Information sharing, Process control, and Partner collaboration [2]. They identified that the technology "enables designing new ways of working, i.e., new business models. The ITAIDE technologies .. facilitate a redesign of the interaction pattern between government and businesses, and second, by facilitating piggybacking as a process optimization." (ibid. p. 179). The idea of piggybacking enable actors involved to reuse electronic information instead of retyping whereby work effort is reduced and the data quality is improved. Several living labs were developed and successfully evaluated with a range of leading IT vendors illustrating the fact that by following the ITAIDE approach, it's possible to make a significant improvement in trade facilitation. In a subsequence research project Cassandra (2010–2014) the electronic data pipeline concept was elaborated for enhancing visibility in international supply chains by an event driven architecture providing more up to date/real time data to improve the logistic efficiency [6, 7]. Based on stakeholder analysis more solutions for business-to-government information have been developed e.g. a customs dashboard [8]. The above mentioned projects have successfully developed local IOS solutions primarily based on EDI (and standards) for demonstration purposes which subsequently not have been adapted by the organizations involved; one of the major reasons given is that the organizations are reluctant to share their information/documents. In summary the idea of an electronic pipeline dates years back and the solutions tend to focus on harmonizing/standardize information. This research contribution is to provide complementary design properties which affect the architecture of the SIP.

2.2 Inter-Organizational System for International Trade

In more general terms, IOS are defined as "information systems to span boundaries between countries, organizations and the relatively separate components of large, geographically dispersed corporations" [9]. Extant literature about using IT for collaboration across organizational boundaries and borders is typically studied under the umbrella of IOS [4]. A closer look on the IOS literature reveals that there are more than 25 theories [10] and no single theory stands out as predominant. The majority of research regarding IOS is focused on EDI[9] [11], and a majority of the described IOS are successfully utilizing EDI [12]. For international trade, the benefits of facilitating IOS based on EDI are well documented [13] but the cost of change is relatively high [14]. The proposed SIP can be categorized as an IOS according the IS theory on IOS [9] since the SIP is an

[6] www.cassandra-project.eu.

[7] www.i-cargo.eu.

[8] "Real-time monitoring is the capability to monitor and log real time – where a shipment is and how it is handled. Process control is the capability to document and evaluate that business processes meet control standards. Information sharing means the ability to electronically exchange information regarding shipments with trading partners and authorities. Partner collaboration refers to the joint capability of trading partners and IT providers to develop end-to-end control and transparency".

[9] Based on the international EDI standard: United Nations/Electronic Data Interchange For Administration, Commerce and Transport (UN/EDIFACT) developed under United Nations.

information systems with the purpose to span boundaries between organizations with separate components/systems and borders of countries which are geographically dispersed. The examples of IOS includes "electronic data interchange .. supply chain management, electronic funds transfer, electronic forms, electronic messaging, and shared databases" [15]. Note that, though it's widely used for the collaboration, e-mails are not considered as IOS because the IOS focuses on system-to-system connection and not system-to-human. Similar the SIP focuses primarily on systems and leaves the actor/user centric element to the individual organizations. The majority of the IOS researched are based on standardized EDI messages and in this regard the design of the SIP is different since the standardization is limited to the identifiers and the few additional data (needed for the subscribe service). So, with respect to standardization, the key focus of the SIP is on standardization of IS systems interfaces, and not on standardization of data and or messages that are exchanged between the systems.

Even IOS based on EDI is relatively successful; it's primarily used by large organizations and only covers a small part of the communication of information and documentation involved in international trade [16]. The widespread use of EDI based IOS [4] are mainly automation islands (mainly due to the standardization required primarily is successful locally), which are not properly integrated especially not internationally, accordingly there is a huge need for some type of inter-organizational reengineering to reduce mistakes, increase efficiencies and reduce time lag. The SIP can reengineer the IOS by offering communication across multiple organizations and borders complementing the existing traditional EDI based IOS. By designing the SIP to a lower cost and entrance investment than the existing IOS the SIP is expected to be used by more organizations and cover more information/documentation.

2.3 Design Theory for Information Infrastructure

The components of an IS design theory includes: (a) Requirements (b) Set of system features (c) Kernel theory and (d) Design principles/properties [17]. Hanseth and Lyytinen [18] propose a design theory for IIs based on IS design theory with a kernel theory and a set of refined properties for II. The emergent properties are: Shared, Open, Heterogeneous, and Evolving. Formulated theoretically, IIs is defined "as a shared, open (and unbounded), heterogeneous and evolving socio-technical system consisting of a set of IT capabilities and their user, operations and design communities." [19]. Additionally are identified two structural properties: Organizing principle and Control. Based on the kernel theory, a set of design properties and nineteen design rules for II has been suggested as design strategy addressing the two generic problems for IIs: bootstrapping and adaptability, bootstrapping being the initial start up and adaptability being the spreading of use.

The proposed SIP can be characterized as an II according to the IS definition of IIs since the SIP complies with both the emergent properties: Shared, Open, Heterogeneous, and Evolving; and structural properties: Organizing principle, and Control. Additionally the SIP is intended to have a global reach, being open to any organization, to be one virtual pipeline build on many pipelines, and to be realized through evolution; accordingly the SIP is regarded as an II. The IS theory to guide the design and development of a successful SIP would then be the design theory for IIs [19], which includes a set of

design properties and design rules for successful bootstrapping and evolution. The organizing principle for the SIP is not settled yet but it's intended to facilitate the emergent properties where the control is distributed to the involved organizations.

Note that Hanseth and Lyytinen [19] do not distinguish between the different types of communication in the II since they have both industry specific EDI based IOS platform and the Internet as examples of II. The concept of communication makes a crucial design property difference between the EDI based IOS and the "Internet" for shipping information. Furthermore the governance of trust and protection of information is found to be a key design property where the SIP differentiates between published shared and trusted information shared only bilaterally after authentication. In above has been discussed to which extent the SIP can be categorized as IOS, a platform and/or an II according to the IS theory. The proposed SIP can be characterized as an II and an IOS, even it's development is very different from the development of standard EDI message based IOS. Accordingly the IS theory is providing guidelines for successful design and development of the SIP.

3 Research Method

The research method applied for the research reported is following an IS design science paradigm, and the interventions are described in the following. The initial focal case for the design of the SIP has been the trade lane transporting fresh cut roses from Kenya to Europe. Design science research is a particular perspective within IS research [20] which focuses on the development of artifacts related to information and communications technology. Design science research includes an evaluation of the designed artifacts. For the initial design the evaluation is artificial and not naturalistic based on real use of the SIP which follows in a later when the prototype is tested on actual shipments. Design science research places IS research between the environment (practitioners) and the knowledge base (researchers), the knowledge justifies the proposed solution to the problem in the environment. In the case of the SIP, the exchange is between a network of organizations within international shipping and the IS research field's knowledge, and the research focuses on relevance and rigor guided by a set of seven "guidelines for Design Science in Information Systems" [21]. This research has been inspired by those guidelines e.g. for the design and evaluation phases.

One of the key features of a design science project is iterations of interventions between practitioners and researchers. The interventions with both researchers and practitioners range from dedicated workshops, meetings and conference calls to workshops over 1–1½ year from spring 2014. The interventions have been documented by written material in the form of minutes of meetings and presentations, which in the subsequence interventions has been commented upon in order to validate the correctness of the documentation.

4 Design Properties for the Shipping Information Pipeline

In this section is presented the key design properties for a domain specific II for international containerized trade bound for European Union. The idea was first conceived in

the IT department of one of the stakeholders (a large international container shipping line), which initiated an initial design named the SIP. Several activities have taken place towards prototyping of the SIP, which involved identifying: the conceptual idea, the business benefits and the associated business model, the major stakeholders, the issues addressed by the potential use, the potential barriers/obstacles, etc. In the following is only the design properties described.

The conceptual idea has been communicated between participants from of the potential stakeholders, who defined the SIP in the following way: "The SIP is a service based facility to allow partners in the supply chain to share accurate original data from it's source. It can connect any number of trading partners .." (minutes of meeting from a workshop 3rd September 2014 between potential stakeholders). They identified two main issues: Lacking full end-to-end supply chain visibility and lacking the ability to efficiently share common data/documents.

Another set of stakeholders focused on other but related main issues: (I) the potential for reducing the relatively high cost of the administrative barriers for international trade; (II) the security challenge for the authorities; (III) lack of visibility in the supply chain for international trade. Different stakeholders focus on different issues to be addressed by the SIP. The potential users have been identified by a mapping actors among the more than thirty organizations involved in one selected trade lane for international trade and encountered actors from both private and public organizations in more than five countries. Additionally the information primarily in the form of documents to be shared were listed and characterized. The businesses benefits for the potential organizations utilizing the SIP have been exemplified in order to identify possible business models for all relevant actors potentially affected by the new SIP. There seems to be a challenge identifying a feasible business model for using the SIP.

The potential stakeholders for the SIP have been analyzed and include e.g. IT vendors and start ups. No stakeholder stands out as being the obvious 'key-stakeholder', who is able to set the standards and enforced it. Every stakeholder holding an installed based have been positive to the idea of communicating using the SIP, but none of them have taken the lead. Instead each and every stakeholder has established collaboration with all the other relevant stakeholders, and consequently more or less everyone is in principle prepared to take part in an overall initiative, but nobody is prepared to make a commitment to lead. It seems to be a paradox that nearly all of the involved organizations are expected to benefit from the SIP but none of them sees SIP as their core business. One of the major stakeholders in the supply chain took initiative to design a first version of the SIP and the key design properties are described in the following. The design of the SIP is kept very clean and with one clear focus: sharing of information about events for shipments relevant for the actors in the supply chain for international trade. For the communication to potential stakeholder the conceptual idea behind the SIP has been formulated in design properties /criteria:

- **No Big Brother.** In order to avoid big brother issues where a central entity has access to central database with detailed information on global trade, no trade document are stored in or transferred via the Shipping Information Pipeline. (partly relates to the structural design properties of II: organization and control)

- **Integrate Once – Connect to Everyone.** When an actor has built the required standardized integrations, the actor will be able to exchange information seamlessly with all other actors integrated with the Shipping Information Pipeline. (relates to the design property of II: shared)
- **One Virtual Pipeline Build by Many Physical Pipelines.** For the users the Shipping Information Pipeline will look like one pipeline, but the actual physical infrastructure can be handled by several individual organizations. Standards will ensure the Shipping Information Pipelines integrates seamlessly in a way similar to how the Internet works today. (relates to the design property of II: heterogeneous)
- **No Facilitation of Commercial Agreements.** The Shipping Information Pipeline will not facilitate commercial agreements between two actors." (is not addressed by any of the design properties of II)

The above design properties/criteria from presentation at a workshop 16th October 2014 reflect the focus on creating trust for the organizations using the SIP, ease of use and global coverage. The trust is addressed by not publishing, sharing or storing detailed data about shipments in the SIP but leaving that to a direct bilateral connection among the ones wanting to share. Additionally excluding commercial agreements increase the trust since no commercial data are in the SIP to prevent any use of the SIP to gain commercial benefits on behalf of competitors are not facilitated by the SIP. When an organization has integrated to the SIP then the exchange of event information will be seamlessly and easy to use e.g. from inside the organizations' IT systems. The global coverage of the SIP is ensured behind the scene even regional set of SIPs are foreseen demanded by practicalities and by authorities.

The above description of the design considerations, including the design properties communicated by the stakeholder taking the initiative, illustrates that the design of the SIP involves many dimensions without an overall consent about the actual properties. The design of the SIP complies with all the design properties from design theory for II and is in line with the ideas from previous research: information from the source via an electronic pipeline, piggy bagging, and up to date/real time information; but to focus only the key design properties were communicated. Further the key design properties communicated changed depending on the actual utility of the particular organization and the extent to which it's possible to address their concerns. The key design properties communicated for the SIP could be characterized as not being a fixed set of properties but rather to be flexible, evolving and adapting to the audience.

5 Evaluation of the Shipping Information Pipeline

The evaluation of the SIP is an ongoing process where various potential stakeholders evaluate the SIP typically at different abstraction levels. The levels include actor level, organizational level, country/region level and society level.

The individual actor in the many organizations using the SIP will be able to get more insight into events in the supply chain for international trade for the shipments in which the individual actor is potentially interested since today none of the actors have transparency.

Additionally the SIP will provide higher quality and up to date information compared to today where information often is missing, out of date and of poor quality information and out of date information which is a major headache for the actors. Accordingly, when asked they find the service provided by the SIP useful especially when things do not go as planned.

The private organizations involved are the traders and the service providers. The traders foresee that the SIP can improve the possibilities for more efficient logistical coordination and lower the risk which will impact the international trade cost and willingness to trade. The international trade cost can be split in a physical transportation cost and an administrative cost of respectively 8 % and 20 % of retail cost [22]. The SIP addresses primarily the administrative border related part of international trade cost which is significant and amounts to approximately 20 % of the retail cost. The service providers e.g. a major shipping line (the main drivers behind the SIP is obviously interested) foresee the main benefit being that lower international trade cost will increase trade volume resulting in more business especially when being a first mover.

The public organizations are active in supporting the prototype to realize the idea (including the two officers from customs in NL and UK, who first came up with the idea of utilizing an electronic data pipeline for international trade). Further the public organizations taking part in the piloted trade lane express their willingness to collaborate regarding the pilot project accordingly they foresee potential improvements for their area of responsibility. Through the SIP the authorities will have the opportunity of getting data directly from the source whereby the quality of the information will increase compared to today, which enable the authorities to improve their risk assessments and the accuracy for the calculation of tariffs etc. but it also imply a change for the authorities' way of working since they need to follow a link to get detailed information instead of receiving it (when requested).

Another set of private organizations involved are IT vendors offering solutions that facilitates information exchange for international trade. They have been positively engaged in collaborating regarding the SIP but none of them have seen a business opportunity, which they want to pursue. Anyhow a major IT vendor has been involved in a series of workshops detailing the architectural design of the SIP and has agreed to invest and build a prototype. But the vendor is still struggling to find an attractive business model, hence it is unclear, who is prepared to fund the further development and operations of the SIP. Governing of the SIP is also a challenge for such a hugely diversified group of organizations involved in the SIP.

At country level the impact of reducing the administrative barriers are estimated to have a significant impact on trade volume which affects the economic positively. The World Economic Forum (WEF) estimates that an improvement to half-way of regional best practice and of global best practice will result in increased Gross Domestic Product (GDP) by respectively 3 % and 5 % [23]. Such improvements are important especially for developing countries e.g. in East Africa (Sub Sahara), where the similar estimate is an increase in GDP by 12 % if applying halfway global best practice. For the first piloted trade lane between East Africa and Europe several association expressed positive expectations about the SIP and committed to be actively involved in the first pilot project lead by TradeMark East Africa[10].

[10] Trade and Markets East Africa is an East African not-for profit Company Limited by Guarantee established in 2010 to support the growth of trade in East Africa.

European Union has a clear an interest in the SIP, since EU is actively involved and is funding the research program in which the testing of the SIP prototype is a part. The EU sponsors especially aim to improve the security for containers imported to the EU and to ease trade between US and EU.

At society level WEF estimates that by lowering barriers for international trade volume will increase and thereby fuel economical growth. "Estimates suggest that an ambitious improvement in two key components of supply chain barriers, border administration and transport and communications infrastructure, with all countries raising their performance halfway to global best practice, would lead to an increase of approximately US$ 2.6 trillion (4.7 %) in global GDP and US$ 1.6 trillion (14.5 %) in global exports. By contrast, the gains available from complete worldwide tariff elimination amount to no more than US$ 400 billion (0.7 %) in global GDP and US$ 1.1 trillion (10.1 %) in global exports" [23]. On basis of the above the SIP can potentially contribute significantly to the growth in the global economy.

In summary: (1) a venture fund of a major IT vendor has decided to fund to built a prototype of the SIP based on the design properties (2) a regional pilot implementation is planned for East Africa (3) EU sponsors the research (4) several organizations involved in international trade have committed to participate in testing with real shipments. The willingness to engage and invest in prototyping and testing of the SIP is taken as a positive evaluation of the design properties for the SIP.

6 Discussion

The overall vision of the SIP has guided the development of the key design properties. The design of previous attempted solutions for collaboration in the supply chain for international trade has been EDI based IOS; the design of the SIP is based and in line with the design properties for II plus additional ones but still in line with the initial ideas: information directly from the source, piggy bagging and real time information. One of the advantages is that an IOS based on II as the SIP is built on top of the internet and it's standards requires less standardization efforts compared to the EDI based IOS. Compared to previously attempted solutions the design knowledge is not the EDI based IOS but Design Theory for II.

Only a set of the design properties namely the key design principles have been communicated (in writing) among the stakeholders. Further focusing on the key design properties contribute to make the design clean and simple. The key design properties have evolved and have been adjusted over time primarily depending on the audience, the above described key design properties express the consensus among organizations involved in the prototype even so minor adjustment of the design properties are to be expected. As described above the key design properties for the SIP add design properties to the ones provided by previous research regarding IOS for international trade. The contribution of this paper is to expand previous research with complementary key design properties.

References

1. Jensen, T., Bjørn-Andersen, N., Vatrapu, R.: Avocados crossing borders: the missing common information infrastructure for international trade. In: Proceedings of the 5th ACM International Conference on Collaboration Across Boundaries: Culture, Distance & Technology. ACM (2014)
2. Tan, Y.H., Bjorn-Andersen, N., Klein, S., Rukanova, B. (eds.): Accelerating Global Supply Chains with IT-innovation: ITAIDE Tools and Methods. Springer, New York (2011)
3. Jensen, T., Tan, Y.-H., Bjørn-Andersen, N.: Unleashing the IT potential in the complex digital business ecosystem of international trade: the case of fresh fruit import to European Union (2014)
4. Kaniadakis, A., Constantinides, P.: Innovating financial information infrastructures: the transition of legacy assets to the securitization market. J. Assoc. Inf. Syst. **15**(4) (2014)
5. Branch, M.A.I.: Report on the investigation of the structural failure of MSC Napoli, English Channel on 18 January 2007. MAIB Report (2008)
6. Klievink, A., et al.: Enhancing visibility in international supply chains: the data pipeline concept. Int. J. Electron. Gov. Res. **8**(4), 2012 (2012)
7. Overbeek, S., Janssen, M., Tan, Y.-H.: An event-driven architecture for integrating information, processes and services in a plastic toys supply chain. Int. J. Coop. Inf. Syst. **21**(04), 343–381 (2012)
8. Klievink, A., Janssen, M., Tan, Y.-H.: A stakeholder analysis of business-to-government information sharing: the governance of a public-private platform. Int. J. Electron. Gov. Res. **8**(4), 2012 (2012)
9. Gregor, S., Johnston, R.B.: Developing an understanding of interorganizational systems: arguments for multi level analysis and structuration theory. In: ECIS 2000 Proceedings, p. 193 (2000)
10. Madlberger, M., Roztocki, N.: Cross-organizational and cross-border IS/IT collaboration: a literature review. In: Proceedings of the Fourteenth Americas Conference on Information Systems, Toronto, ON, Canada (2008)
11. Reimers, K., Johnston, R.B., Klein, S.: The shaping of inter-organisational information systems: Main design considerations of an international comparative research project. In: Tan, Y., Vogel, D., Gricar, J. Lenarts, G. (eds.) 17th Bled eCommerce Conference: "eGlobal", Faculty of Organizational Science, University of Maribor, Bled, Slovenia (2004)
12. Robey, D., Im, G., Wareham, J.D.: Theoretical foundations of empirical research on interorganizational systems: assessing past contributions and guiding future directions. J. Assoc. Inf. Syst. **9**(9) (2008)
13. King, J.L.: Balance of trade in the marketplace of ideas. J. Assoc. Inf. Syst. **14**(4), 3 (2013)
14. Henningsson, S., Bjørn-Andersen, N.: Exporting e-Customs to developing countries: a semiotic perspective. In: Proceedings of the Second Annual SIG GlobDev Workshop, Phoenix, USA. Idea Group Publishing, Hershey (2009)
15. Schwens, C., et al.: International entrepreneurship: a meta-analysis. In Academy of Management Proceedings. Academy of Management (2011)
16. Jensen, T., Vatrapu, R.: Ships & roses: a revelatory case study of affordances in international trade. In: 23rd European Conference on Information Systems (ECIS) 2015
17. Walls, J.G., Widmeyer, G.R., El Sawy, O.A.: Building an information system design theory for vigilant EIS. Inf. Syst. Res. **3**(1), 36–59 (1992)
18. Hanseth, O., Lyytinen, K.: Theorizing about the design of Information Infrastructures: design kernel theories and principles (2004)

19. Hanseth, O., Lyytinen, K.: Design theory for dynamic complexity in information infrastructures: the case of building internet. J. Inf. Tech. **25**(1), 1–19 (2010)
20. Gregor, S., Hevner A.R.: Introduction to the special issue on design science. Inf. Syst. e-Business Manag. **9**(1) (2011)
21. Hevner, A.R., et al.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
22. Anderson, J.E., Van Wincoop, E.: Trade costs. National Bureau of Economic Research (2004)
23. WEF, W.E.F.i.c.w.T.B.C.G., Connected World. Transforming Travel, Transportation and Supply Chains. World Economic Forum, Insight Report, 2013

Enhancing Awareness on the Benefits of Supply Chain Visibility Through Serious Gaming

Tijmen Joppe Muller[1(✉)], Rainer Müller[2], Katja Zedel[2], Gerwin Zomer[1], and Marcus Engler[2]

[1] TNO, PO Box 23, 3769 DE Soesterberg, The Netherlands
{tijmen.muller,gerwin.zomer}@tno.nl
[2] Institut für Seeverkehrswirtschaft und Logistik, Universitätsallee 11-13, 28359 Bremen, Germany
{rmueller,zedel,engler}@isl.org

Abstract. Improving both efficiency and security in international supply chains requires a new approach in data sharing and control measures. Instead of managing supply chain risks individually, supply chain partners need to collaborate in order to exchange cargo information and implement control measures on the level of the entire supply chain. Governmental agencies, having access to this up-to-date and complete information, can implement alternative risk assessment policies, resulting in less disruptive ways of supervising entire trade lanes. However, this paradigm shift requires awareness of these supply chain visibility concepts and increased collaboration between partners in a value chain. In order to disseminate these new concepts and initiate cooperation between key stakeholders, a serious game called 'The Chain Game' was designed, implemented and evaluated.

1 Supply Chain Visibility

One of the main challenges that international intermodal container logistics faces today is how to balance efficiency and security. Efficiency is one of the key performance indicators businesses are aiming for in order to have competitive advantages. In contrast, governmental agencies, such as customs at border control, have to maintain a high level of security to protect citizens and business partners against unlawful practices. However, increased efficiency and increased security are goals that may oppose each other. For example: in order to reach maximum efficiency, disruptive governmental interventions are undesirable from the business perspective, whereas maximum security is reached by high inspection rates and heavy enforcement measures applied to all incoming and outgoing containers. In current day practice, most organizations in a value chain manage their supply chain risks individually, by applying a combination of the 4Ts: transfer, terminate, tolerate and treat. Risk transfer practices are very mature and widely applied by the more powerful chain actors. The weaker chain actors often cannot transfer such risks and pressure on margins forces them to tolerate certain risks. If risk treatment is being considered, it is often done through internal control measures.

M. Janssen et al. (Eds.): I3E 2015, LNCS 9373, pp. 503–512, 2015.
DOI: 10.1007/978-3-319-25013-7_41

Research in supply chain dependencies, vulnerabilities and resilience proposes an alternative approach, one that considers the risks from the perspective of the entire supply chain, instead of from each individual fragment of the chain. Risk based control and supervision seems the way forward, targeting high risk consignments and concentrate inspection effort on the targeted categories, whilst applying alternative less disruptive ways of supervision to the rest. This concept of *Supply Chain Visibility* [1] takes the needs of both governmental agencies and businesses in the chain into account. The basic idea is a data sharing concept that is built on integration of existing data processing systems. This so-called 'Data Pipeline' is used by business companies for end-to-end supply chain visibility, to ensure high data quality and data completeness and efficient connectivity to community systems. Customs and other authorities can connect to the Data Pipeline through a dashboard interface and use selected data for reporting purposes, such as customs declarations. It also allows them to piggy back on the validated and enriched supply chain data and advance the risk analysis and assessment they used to perform on data from declarations. An advantage for all parties is that all available information is up-to-date, since the data itself is fed by the originator of data and only retrieved by authorized users. Time-consuming search for additional information concerning a single transport could be avoided and additional data sources (from container security devices for example) could enrich the data sets without revision of company-wide or governmental data bases.

In theory this Supply Chain Visibility concept can benefit all actors within a supply chain. This approach may not only lower costs for the business partners in the supply chain because of increased data sharing efficiency, but may also have customs recognize that the corresponding trade lane partners are 'in control' of the major risks in their supply chain, resulting in less disruptive interventions from the authorities. However, some required steps need to be taken in practice to operationalize this approach. Most importantly, trade lane partners need to collaborate in implementing both the data sharing concept and additional chain control mechanisms. Examples of control mechanisms on the level of the entire chain are partner screening and coaching, applying technologies for container integrity (e.g. Container Security Devices) and data validation procedures, either through physical checks (e.g. a tallyman checking the packing list information) or by comparing data from different sources (three-way match). A second important step is new policy from the governmental agencies, that allows less disruptive interventions based on a different way of risk analysis.

Creating awareness of the aforementioned concepts and control mechanisms, having key stakeholders in a value chain recognize the benefits of such an approach and initiating collaboration to put these concepts into practice proves to be a difficult and time consuming process. In order to speed up the concept development cycle but also to disseminate the research findings we implemented an alternative approach: a serious game. We believe the characteristics of serious gaming, which allows stakeholders to have concrete experiences with the innovative concepts in interaction with each other, will disseminate the research successfully and will initiate collaboration for future implementation of supply chain visibility.

2 Serious Gaming

Serious games have proven to be an excellent medium to explain and teach complex concepts through experiential learning in a wide variety of domains. A well-known example is strategic war games that have been used in military training for centuries. At least since the mid-19th century the use in the military tactical training of officers has been documented [2, 3]. Business simulations, which can be simple board games or massive computer games, are commonly used in management education [4, 5]. An almost classic example for this is the Beer Game, which was created in early 1960s at the MIT's Sloan School of Management [6] and is still used in management education to have players experience the bullwhip effect. Technological progress has allowed for more complex and realistic simulators, such as flight simulators, nautical simulators, crane simulators or driving simulators.

Simulation games and simulators are commonly used to allow students to experience situations which are too costly, hazardous or unethical to experience in the real world. Simulation games are very useful to understand complex problems through experiential learning: experiencing the effects and dependencies of own behavior is a very strong learning method. Simulation games allow making mistakes in a safe environment and thus learning by trial and error [7]. This invites players to try new strategies and challenge long-standing patterns of behavior. In "The Chain Game" we use this to guide the players to try more collaboration.

The use of (computer) games in education - either especially designed for education or commercial off-the-shelf (COTS) games - is growing [8–10]. The learning theories which can be identified in games include well known theories of learning such as Gagne's "Conditions of Learning theory", Gardner's "Theory of Multiple Intelligences", Skinner's "Operant Conditioning theory", Thorndike's "Laws of Effect", Maslow's "Hierarchy of Needs theory" and Kolb's "Experiential Learning Model". Also incorporated are learning approaches like active learning, experiential learning and situated learning [11].

Probably the most compelling point about games is their ability to motivate and engage, far better than other teaching methods. Prensky believes that games "are the most engaging intellectual past-time that we have invented" [8]. In studies, participants report to have more interest in simulations or games than in classroom instructions. This correlates with the participants investing significantly more time in learning with games [7, 9, 11]. For our dissemination we take advantage of the engaging and interest raising nature of games. Buckley and Anderson identified three common characteristics which make games so compelling [13]:

1. Being in control: Players work at their own ability level and speed and repeat material as needed.
2. Feedback and rewards: Games give immediate feedback. Additional most games reward behaviors and actions by the player which are "positive" according to the in-game logic particularly. This increases the frequency of this behavior and teaches a positive attitude towards the content of the game.
3. Challenge and mastery: Games challenge players but remain doable. They keep the balance between too hard and tedium, giving the player feelings of self-efficacy.

Since games are an engaging medium for explaining complex systems, they can be very useful to create awareness for new concepts, such as the data sharing concept of Cassandra, with key stakeholders. Serious gaming allows meaningful interaction with both a relevant representation of the world and between key stakeholders, allowing them to analyze the situation from different perspectives and create a shared understanding of the challenges [14]. As such, gaming can play an important role in the transformation process needed to eventually implement these concepts.

3 The Chain Game

Given the goals we hope to achieve, explaining the complex and abstract concepts of supply chain visibility and initiating collaboration between key stakeholders in supply chains, we believe serious gaming is a very suitable medium – more so than reports or presentation material. In modern supply chain logistics stakeholders face multidimensional problems with complex side effects and non-linear dependencies. A game typically is a simplified world, reducing this complex multidimensional problem to its essentials in order to showcase only the relevant parts. This section describes *The Chain Game*, the serious game that was designed to transfer the abstract chain visibility and chain control concepts into practical examples and to allow players to experience the effects of the researched innovations, even though the exact effects (in reality) are not yet known.

The starting point for designing The Chain Game is the goals that should be achieved after playing the game. These learning goals were defined as follows:

1. Players understand that collaboration between supply chain partners is needed in order to operationalize the concept of supply chain visibility. Additionally, collaboration will have benefits on the performance of the supply chain, but the investing partners will not necessary receive the gains, so discussion on return-on-investments is necessary. In general, the performances of businesses within one chain are interdependable.
2. The concept of supply chain visibility and the underlying innovations need to be disseminated. The players (typically key stakeholders in a supply chain) need to understand the broader picture of supply chain visibility, but also understand typical control measures they could implement.
3. Implementation of the supply chain visibility concepts will increase robustness of the entire supply chain as risks are more easily mitigated and governmental agencies will apply less disruptive supervision measures.

These requirements are used to design components such as the world model (cause and effect), physical setting, narrative (i.e. scenario and storyline), aesthetics, but also the game session (i.e. the process of game play). This design is an iterative process in which play-testing is key: by trying out the game, the designers can observe whether the goals are reached and interaction and user experience are what they hoped for. In the sections below, we explain how the design meets the initial requirements.

3.1 Benefits of Collaboration

In order to show the effects of collaboration and initiate discussion between key stakeholders, we needed multi-player gameplay. The Chain Game has a fixed amount of five roles, of which each plays part of an international supply chain: two sellers, two freight forwarders and one buyer. The goal within the game is to create highest value for the own company, based on three key performance indicators (KPIs). In order to do so, players can invest in innovations from their limited resources. Innovations are either individual (conform the internal control measures) or collaborative (in which case collaboration obviously is necessary) and will improve specific KPIs of specific companies in the chain.

Interdepence between players is designed in the game model in multiple ways. First off, investing in collaborative innovations will increase chain value more than individual innovations on the long term. However, the collaborative innovations do not necessarily benefit the players that have invested in them: they may benefit one player more than the other, may benefit a player that has not invested at all or may even decrease company value of an investing player. These effects can be reduced by the players themselves, as they are allowed to share their benefits with each other. This game mechanic stimulates investing in collaborative innovations, but also supports discussion on return-on-investment. Additionally, the fact that the players are (collectively) responsible for implementing innovations and they almost immediately see the effects of their actions contributes to the game characteristics 'Being in control' and 'Feedback and Reward' as described by [13].

Second, company value is not only based on the key performance indicators of the *own* company, but also on the performance of the chain as a whole. But the performance of the chain depends for a large part on the weakest link of that chain. As a result, players need to convince other players to improve certain KPIs, conveying the message that chain partners are highly interdependent.

3.2 Disseminating Concepts

The supply chain innovations are made available as *actions* to the players of the game. This forces the players of The Chain Game to become familiar with the concept of Supply Chain Visibility: they are required to think about the value of the innovations in the context of their working environment and will experience the necessity to cooperate in order to create greater supply chain visibility. The innovations are focused on custom-related interruptions, in line with the research goals of the Cassandra project, and described on an abstract level, in such a way that they have an effect on one or more KPIs of one or more roles. The effects of an innovation are not disclosed until the innovation is implemented, creating the challenge for the players to understand these innovations and estimate their effects if they want to master the game (as described before [13]). As they are trying to implement the best possible innovation for their company, the main goals of the game are achieved: communicating the concept of supply chain visibility, understanding of the underlying innovations and discussion within the supply chain on these innovations.

An example of a supply chain visibility concept is 'Exception Reporting', which is made available as collaborative innovation to the players of The Chain Game (see Fig. 1). For 'Exception Reporting' three investors are required: both freight forwarders are forced to invest, and a third one (either one of the sellers or the buyer) also needs to contribute to implement this innovation. As described before, the companies that invested in the innovation are not necessarily the companies that receive the (dis)advantages: 'Exception Reporting' not only improves the KPIs for both investing freight forwarders, but also the buyer, whether he invested or not. An important step in clarifying the concepts to the players is explaining the argumentation of the effects, which is done by a game facilitator. In the case of the 'Exception Reporting' innovation, the reason for the impact is described as follows: "Exception Reporting provides the Freight Forwarders with early warnings for expected errors, allowing them to take measures in an early stage. This will give them more control over their business and they will have the ability to reduce the negative consequences, which is also of interest for the Buyer."

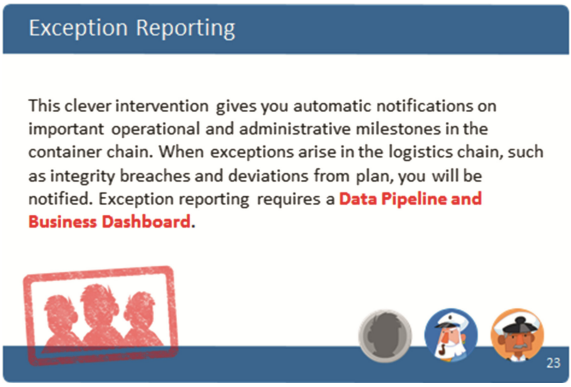


Fig. 1. An example of a chain innovation card

A big challenge of designing The Chain Game was that insufficient data is available on the exact effects of these innovative concepts, so it was not possible at this point to create a simulation model that predicts the effects on a detailed level. However, the innovations were already described on an abstract level in the CASSANDRA Project deliverables [15]. As the goal of the game is to create awareness about this knowledge, a qualitative model suffices to create this awareness and support discussion about the various concepts. As a result, the effects of each innovation were established by interviewing subject matter experts in a series of workshops to come to the content model of The Chain Game. In total, the effects of 6 individual innovations and 26 collaborative innovations were estimated and implemented this way.

3.3 Chain Robustness

One of the promises of increased supply chain visibility is that the supply chain will be less sensitive to incidents that may occur. To communicate this message, incidents occur

regularly to test the current state of the supply chain. Incidents typically have a negative effect on one or more KPIs of one or more companies, but implemented innovations may (partly) mitigate these effects. The effects of an incident and the mitigating innovations for each incident have been established by the subject matter experts, similar to the effects of the innovations. An example of an incident is 'Cargo Theft': "Cargo has been stolen from trucks that were on their way from the Seller to the seaport. Trucks are totally emptied by this criminal organization." It has a negative effect on the freight forwarders, the seller and the buyer with the following argumentation: "Both of the Freight Forwarders are held responsible for their lack of security measures for their part of transportation. Their reliability is severely damaged and they have to compensate the costs of the theft and therefore their capability weakens. The Buyer doesn't get its raw product in time, which means that yet again, there will not be enough of his product available for retailers. This hurts the perceived reliability of the Buyer. The Seller is afraid to lose clients and quickly offered to deliver new cargo at cost price." The innovation 'Exception Reporting' discussed before mitigates 25 % of these effects.

A second game mechanic that gives players the feedback on the state of their supply chain is the role of Customs. Customs is strongly interested in supply chain visibility, since they profit from additional information in a way that checks on data might be executed in less time. Therefore, the partners in the supply chain may gain advantages in relation to Custom checks if they provide a certain degree of visibility. Ideally, Customs perceives the supply chain as a *trusted trade lane*, a concept strongly related to supply chain visibility. These dynamics reflect in the game by having Customs to perform a risk assessment on the state of the supply chain at the end of each game round. The risk assessment decides if a physical check of the goods is performed. Such a check has negative effects on the reliability of all stakeholders of the entire supply chain, since goods transported may be delayed for days. However, a number of innovations in the game that are of interest to Customs will have the supply chain proceed towards the certificate of trusted trade lane. Each implementation of one of these innovations decreases the probability of a physical check.

4 Evaluation

The final version of The Chain Game was disseminated by playing it with two groups of stakeholders: consultants of a large accountancy firm and managers of Dutch Customs. The goal of the play sessions was to both evaluate and disseminate the game with stakeholders. The game was evaluated by a questionnaire; additionally, observations during the reflection phase are also used for the evaluation. A play session consisted of 20 min of introduction, 10 game rounds (75 min), filling in the questionnaire and 30 min of reflection. The goal of the evaluation was to assess whether the game is valuable for creating supply chain visibility awareness and to make an inventory of possible opportunities to apply the game. The questionnaire consisted of 3 questions (see Fig. 2) that needed to be scored on a 5-point Likert scale (1 = *totally disagree* to 5 = *totally agree*), and the following open questions:

- What part of The Chain Game should stay the same?
- What part of The Chain Game should be adjusted?

- Which (business) opportunities do you see for this game?
- Any further remarks?

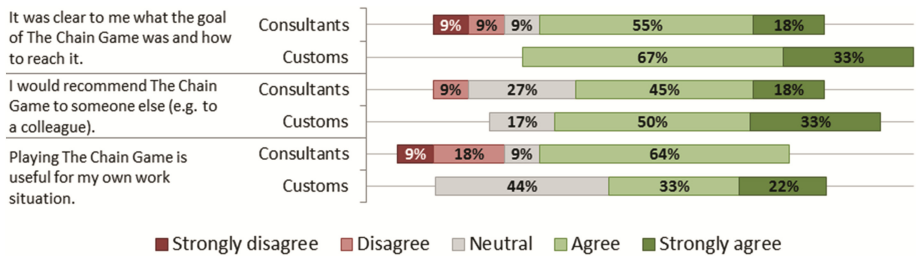


Fig. 2. Results of the evaluation questionnaire

4.1 Results

In general it can be concluded that both groups are convinced The Chain Game has value in creating awareness for supply chain visibility. Figure 2 on the next page shows the results of the three closed questions – the results of question 3 (as shown in Fig. 2) show that both groups are fairly positive about using The Chain Game in their own working environment: the 64 % of the consultants are positive (27 % negative) and 55 % of the Customs group are positive (0 % negative). We believe these score may be toned down somewhat by the fact that these groups were not the primary audience for The Chain Game. For example, Customs does not have an active playing role in the game and for consultants the focus on customs-related interruptions covers only a part of their work. On the other hand, socially desirable answers may also have influenced the results, as the questionnaire was hardly anonymous.

The answers to the open questions and the observations of the game facilitators show that both groups are positive about the way of using a game to raise awareness and understanding for complex information. This was also easily observed: players were drawn into the game from the first round on and the game could keep the attention until the final round, for more than an hour. Additionally, the interactivity and direct feedback was praised by both groups, allowing players to quickly gain insight in the effects of certain innovations. Finally, from the group of consultants several people mentioned they were drawn to the digital and visual implementation.

Both groups believed that more time should be reserved for the facilitator to explain the effects of innovations. Sometimes, the effects of innovations or incidents were challenged by a player, but this does not really stand in the way of the goal of the game: creating awareness. Discussion about the effects in real-life situations are a consequence of the fact that the model is an estimate of subject matter experts and not evidence-based. This means innovations sometimes had effects that the players did not expect. This cannot be resolved in the game, as effects need to be undisclosed to force players to think about the meaning of innovations. However, it is necessary for the game facilitator to take sufficient time for the argumentation *behind* the effects that take place. An expert

opinion supports relevant discussion and counters players' frustration. Additionally, some game design choices have resulted in the omission of factors that can be of importance in the real world. For example, the number of players is fixed, in order to keep the game economy as simple as possible (i.e. a player can basically invest only in one innovation each turn). As a result, stakeholders cannot enter or leave a logistics chain, an action that can have big effects in the real world. This pragmatic choice was made to keep the game playable within a few hours.

Somewhat related is the *game process*: in the play sessions, hardly any reflection on the adopted strategy was possible during the playing of the game because of the fast pace during these sessions. As a result, players were not able to reflect upon the played innovations, the expected and the real effects and the underlying model. This prevented them in creating a broader understanding and forming and adjusting their strategy. Particularly the group of consultants requested one or two reflection phases during the game, in order to improve understanding, which can easily be adjusted by pausing the game after 3 or 4 rounds.

Interestingly, the group of consultants believed the game to be strongly focused on Customs-related innovations, while the group Customs would like a more active role for Customs in the game. The game can be improved for the consultants by adding taxes- and legislation-specific innovations; this is no problem from the perspective of the game design, as long as the effects are sufficiently realistic and in balance with each other. However, it will change the focus of the game, which was primarily aimed at Customs-related interruptions. Making Customs interactive in the game is more difficult to implement, since this would mean adding a new role that is even more asymmetrical to the roles that already exist in the game.

A number of different **opportunities** were put forward during the play sessions. The group of consultants believed the game can be played with companies to raise awareness about supply chain visibility and on the impact Customs can have on supply chains. Both groups believed the game can provide insight in the interdependence between supply chain partners and the importance of jointly investing in innovations and prevent weak links in the chain. Finally, Customs believed the game can be used as learning tool in management studies related to logistics.

5 Conclusions

Disseminating the concepts and ideas concerning supply chain visibility is a challenge, as they are abstract and hard to grasp. Additionally, future implementation of these concepts require both support and collaboration of a multitude of stakeholders, hence awareness within this group of stakeholders is essential to effectuate innovations. In order to explain the complex concepts and initiate collaboration between stakeholders, a serious game named The Chain Game was developed. The goal of The Chain Game is to have players experience the effects and advantages of increased supply chain visibility. The serious game focuses on a variety of innovative concepts, which can be implemented by the main actors in a supply chain. Feedback consists of the effects of

these innovations on a supply chain level. A qualitative game model was built by a team of game designers and subject matter experts, as little evidence-based data is available as of yet.

The fully digital serious game has been tested by stakeholders in two pilots. Both groups evaluated the game as valuable in raising awareness on the subject of supply chain visibility with stakeholders and believed it can act as a supporting tool in bringing the necessary parties together. Additionally, the game can be used as learning tool in management studies related to logistics. An important observation during the tests is the crucial role of the facilitator as subject-matter expert: he needs to have the time and the knowledge to explain the argumentation behind the presented effects during the game. If this precondition is met, The Chain Game can provide an excellent starting point for further discussion on the future of supply chain visibility.

References

1. The CASSANDRA project (2014). http://www.cassandra-project.eu/
2. von Hilgers, P.: Eine Anleitung zur Anleitung. Das taktische Kriegsspiel 1812–1824. Int. J. Study Board Games cnws 200059 (2000)
3. Pias, C.: Computer Spiel Welten. (Diss) (2002)
4. Keys, B., Wolfe, J.: The role of management games and simulations in education and research. J. Manag. **16**(2), 307–336 (1990)
5. Tanner, M., Lindquist, T.: Using MONOPOLY and teams-games-tournaments in accounting education: a cooperative learning teaching resource. Acc. Educ. **7**(2), 139–162 (1998)
6. Forrester, J.W.: Industrial Dynamics. Mass, Cambridge (1961)
7. Randel, J.M., Morris, B.A., Wetzle, C.D., Whitehead, B.V.: The effectiveness of games for educational purposes: a review of recent research. Simul. Gaming **23**, 261–276 (1992)
8. Prensky, M.: Digital Game-Based Learning. McGraw-Hill Companies, New York (2001)
9. Fletcher, J.D., Tobias, S.: Using computer games and simulations for instruction: a research review. In Proceedings of the New Learning Technologies 2006 Conference, Warrenton, VA: Society for Applied Learning Technology (2006)
10. Charsky, D., Mims, C.: Integrating commercial off-the-shelf video games into school curriculums. TechTrends **52**(5), 38–44 (2008)
11. Tang, S., Hanneghan, M., El Rhalibi, A.: Introduction to games based learning. Games Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices, pp. 1–18. IGI Global, New York (2009)
12. Foreman, Joel, et al.: Game-based learning: how to delight and instruct in the 21st century. Educause Rev. **39**, 50–67 (2004)
13. Buckley, K.E., Anderson, C.A.: A theoretical model of the effects and consequences of playing video games. Playing Video Games: Motives, Responses, and Consequences, pp. 363–378. Lawrence Erlbaum, London (2006)
14. Janssen, M., Klievink, B.: Gaming and simulation for transforming and reengineering government: towards a research agenda. Transform. Gov. People Process Policy **4**, 132–137 (2010)
15. Zomer, G., et al.: Deliverable D9.1 CASSANDRA Final Report. Delft, pp. 9–24 (2014)

Author Index