Peter Parycek · Olivier Glassey
Marijn Janssen · Hans Jochen Scholl
Efthimios Tambouris
Evangelos Kalampokis · Shefali Virkar (Eds.)

LNCS 11020

Electronic Government

17th IFIP WG 8.5 International Conference, EGOV 2018
Krems, Austria, September 3–5, 2018
Proceedings

Springer

Lecture Notes in Computer Science 11020

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/7409

Peter Parycek · Olivier Glassey
Marijn Janssen · Hans Jochen Scholl
Efthimios Tambouris
Evangelos Kalampokis · Shefali Virkar (Eds.)

Electronic Government

17th IFIP WG 8.5 International Conference, EGOV 2018
Krems, Austria, September 3–5, 2018
Proceedings

Springer

Editors
Peter Parycek
Danube University Krems
Krems
Austria

Olivier Glassey
University of Lausanne
Lausanne
Switzerland

Marijn Janssen
Delft University of Technology
Delft
The Netherlands

Hans Jochen Scholl
University of Washington
Seattle, WA
USA

Efthimios Tambouris
University of Macedonia
Thessaloniki
Greece

Evangelos Kalampokis
University of Macedonia
Thessaloniki
Greece

Shefali Virkar
Danube University Krems
Krems
Austria

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-98689-0 ISBN 978-3-319-98690-6 (eBook)
https://doi.org/10.1007/978-3-319-98690-6

Library of Congress Control Number: 2018950650

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This year's conference was held after the merging of the IFIP EGOV-EPART conference series with the CeDEM conference. This successful merger resulted in the IFIP EGOV-CeDEM-ePart 2018 conference, which was a high-caliber multi-track conference including a practitioners' track and doctoral colloquium. The conference is dedicated to the broader area of electronic government, open government, smart governance, e-democracy, policy informatics, and electronic participation. Scholars from around the world have attended this premier academic forum for a long time, which has given EGOV a worldwide reputation as one of the top two conferences in the research domains of electronic, open, and smart government as well as electronic participation.

The call for papers attracted completed research papers, work-in-progress papers on ongoing research (including doctoral papers), project and case descriptions, as well as workshop and panel proposals. This conference of five partially intersecting tracks presents advances in the socio-technological domain of the public sphere demonstrating cutting-edge concepts, methods, and styles of investigation by multiple disciplines. The papers were distributed over the following tracks:

- General E-Government and Open Government Track
- General E-Democracy and E-participation Track
- Smart Government Track
- AI, Data Analytics, and Automated Decision-Making Track
- Digital Collaboration and Social Media Track
- Policy Modelling and Policy Informatics Track
- Social Innovation Track
- Open Data, Linked Data, Semantic Web Track
- Practitioners' Track

As in the previous years and per the recommendation of the Paper Awards Committee under the leadership of Olivier Glassey of the University of Lausanne, Switzerland, the IFIP EGOV-CeDEM-ePart 2018 Conference Organizing Committee again granted outstanding paper awards in three distinct categories:

- The most interdisciplinary and innovative research contribution
- The most compelling critical research reflection
- The most promising practical concept

The winners in each category were announced in the award ceremony at the conference dinner, which has always been a highlight of the conferences.

Many people make large events like this conference happen. We thank the over 100 members of the Program Committee and dozens of additional reviewers for their great efforts in reviewing the submitted papers. We would like to express our gratitude to

Noella Edelman, Shefali Virkar, and the team from Danube University for the organization and the management of all the details.

The Danube University Krems is the leading university of continuing education. As the only public university for continuing education in the German-speaking countries, the Danube University Krems sets the standards for lifelong learning. When it first opened its doors to students in 1995, a competence center for scientific specialization was created that focused on the pressing challenges of our times, and whose courses of study are continuously evolving. Today, three faculties with 15 departments are successfully engaged in teaching and research; approximately 18,000 people have already graduated from the University of Continuing Education.

September 2018

Peter Parycek
Olivier Glassey
Marijn Janssen
Hans Jochen Scholl
Efthimios Tambouris
Evangelos Kalampokis
Shefali Virkar

Organization

Electronic Government

17th IFIP WG 8.5 International Conference, EGOV 2018
EGOV-CeDEM-ePart 2018
Austria, Krems, September 3–5, 2018
Proceedings
Part of the *Lecture Notes in Computer Science* book series (LNCS, volume 11020)

Editors

Peter Parycek	Fraunhofer Fokus, Germany and Danube-University Krems, Austria
Olivier Glassey	University of Lausanne, Switzerland
Marijn Janssen	Delft University of Technology, The Netherlands
Hans Jochen Scholl	University of Washington, USA
Efthimios Tambouris	University of Macedonia, Greece
Evangelos Kalampokis	University of Macedonia, Greece
Shefali Virkar	Danube University Krems, Austria

IFIP Working Group 8.5 Elected Officers

Chair

Marijn Janssen	Delft University of Technology, The Netherlands

Past Chair

Hans Jochen Scholl	University of Washington, USA

First Vice Chair

Olivier Glassey	University of Lausanne, Switzerland

Second Vice Chair

Peter Parycek	Fraunhofer Fokus, Germany and Danube University Krems, Austria

Secretary

Efthimios Tambouris	University of Macedonia, Greece

Conference Chairs

Peter Parycek	Fraunhofer Fokus, Germany and Danube University Krems, Austria
Olivier Glassey	University of Lausanne, Switzerland
Marijn Janssen	Delft University of Technology, The Netherlands
Hans Jochen Scholl	University of Washington, USA
Efthimios Tambouris	University of Macedonia, Greece

Local Chair

Noella Edelmann	Danube University Krems, Austria

Track Chairs

General E-Government & Open Government Track

Marijn Janssen	Delft University of Technology, The Netherlands
Reinhard Riedl	Bern University of Applied Sciences, Switzerland
Hans Jochen Scholl	University of Washington, USA

General E-Democracy & e-Participation Track

Peter Parycek	Fraunhofer Fokus, Germany and Danube University Krems, Austria
Efthimios Tambouris	University of Macedonia, Greece
Robert Krimmer	Tallinn University of Technology, Estonia

Smart Governance (Government, Cities, and Regions) Track

Olivier Glassey	University of Lausanne, Switzerland
Karin Axelsson	Linköping University, Sweden
Morten Meyerhoff Nielsen	Tallinn University of Technology, Estonia
Manuel Pedro Rodríguez Bolívar	University of Granada, Spain

AI, Data Analytics, & Automated Decision-Making Track

Evangelos Kalampokis	University of Macedonia, Greece
Habin Lee	Brunel University London, UK
Vasilis Peristeras	International Hellenic University, Greece

Digital Collaboration & Social Media Track

Panos Panagiotopoulos	Queen Mary University of London, UK
Mauri Kaipainen	Södertörn University, Sweden
Ida Lindgren	Linköping University, Sweden

Policy Modeling & Policy Informatics Track

Yannis Charalabidis	University of the Aegean, Greece
Theresa A. Pardo	University at Albany, State University of New York, USA
Noella Edelmann	Danube University Krems, Austria

Social Innovation Track

Gianluca Misuraca	European Commission, Spain
Christopher Tucci	École Polytechnique Fédérale de Lausanne, College of Management of Technology, Switzerland

Open Data, Linked Data, & Semantic Web Track

Efthimios Tambouris	University of Macedonia, Greece
Anneke Zuiderwijk-van Eijk	Delft University of Technology, The Netherlands
J. Ramon Gil-Garcia	University at Albany, State University of New York, USA

Practitioners' Track

Francesco Molinari	Politecnico di Milano, Italy
Peter Reichstädter	Austrian Parliament, Austria

Projects/Reflections & Viewpoints/Workshops/Panels

Marijn Janssen	Delft University of Technology, The Netherlands
Hans Jochen Scholl	University of Washington, USA
Peter Parycek	Fraunhofer Fokus, Germany and Danube University Krems, Austria
Olivier Glassey	University of Lausanne, Switzerland
Efthimios Tambouris	University of Macedonia, Greece

Posters

Thomas Lampoltshammer	Danube University Krems, Austria

PhD Colloquium

Anneke Zuiderwijk-van Eijk	Delft University of Technology, The Netherlands
Gabriela Viale Pereira	Danube University Krems, Austria
Ida Lindgren	Linköping University, Sweden
J. Ramon Gil-Garcia	University at Albany, State University of New York, USA

Senior Faculty

Marijn Janssen	Delft University of Technology, The Netherlands
Peter Parycek	Fraunhofer Fokus, Germany and Danube University Krems, Austria
Hans Jochen Scholl	University of Washington, USA
Efthimios Tambouris	University of Macedonia, Greece

Program Committee

Adegboyega Ojo	NUI Galway, Ireland
Ali Pirannejad	Delft University of Technology, The Netherlands
András Micsik	MTA SZTAKI, Hungary
Annelie Ekelin	Linnaeus University, Sweden
Azi Lev-On	Ariel University, Israel
Barbara Re	University of Camerino, Italy
Bojan Cestnik	Jozef Stefan Institute and Temida, Slovenia
Bram Klievink	Delft University of Technology, The Netherlands
Carl Erik Moe	University of Agder, Norway
Carolin Schröder	TU Berlin, Germany
Chien-Chih Yu	National Cheng Chi University, Taiwan
Christian Rupp	Plenipotentiary for Digitalization, Austria
Cristiano Maciel	Universidade Federal de Mato Grosso, Brazil
Delfina Soares	United Nations University, Portugal
Dmitrii Trutnev	ITMO University, Russia
Dimitris Gouscos	National and Kapodistrian University of Athens, Greece
Divya Kirti Gupta	Hyderabad Business School, GITAM University, India
Elin Wihlborg	Linköping University, Sweden
Ella Taylor-Smith	Edinburgh Napier University, UK
Elsa Estevez	Universidad Nacional del Sur, Argentina
Erich Schweighofer	University of Vienna, Austria
Euripidis Loukis	University of the Aegean, Greece
Evika Karamagioli	Ethniko Kapodistriako University of Athens, Greece
Francisco Falcone	UPNA, Spain
Günther Schefbeck	Austrian Parliamentary Administration, Austria
Harekrishna Misra	Institute of Rural Management Anand, India
Harrie Scholtens	European Institute of Public Administration, The Netherlands
Helen Liu	The University of Hong Kong, Hong Kong, SAR China
Ismael Peña-López	Universitat Oberta de Catalunya, Spain
Jakob Svensson	Malmö University, Sweden
Jesus Cano	IEEE CS eGovernment, King Juan Carlos University, Spain
Joern von Lucke	Zeppelin Universität, Germany
Jolien Ubacht	Delft University of Technology, The Netherlands

Joris Hulstijn	Tilburg University, The Netherlands
José María Moreno-Jiménez	University of Zaragoza, Spain
Judith Schossboeck	Danube University Krems, Austria and City University Hong Kong, Hong Kong, SAR China
Karine Nahon	Interdisciplinary Center Herzliya, Israel
Lasse Berntzen	University College of Southeast Norway, Norway
Leif Sundberg	Mid Sweden University, Sweden
Lotta Groning	Linköping University, Sweden
Luis Luna-Reyes	University at Albany, USA
Luis Terán	University of Fribourg, Switzerland
Luiz Antonio Joia	Getulio Vargas Foundation, Brazil
Luiz Paulo Silva	UNIRIO, Brazil
Maria Nikolova	New Bulgarian University, Bulgaria
Martijn Hartog	The Hague University of Applied Sciences, The Netherlands
Mauricio Solar	Universidad Técnica Federico Santa Maria, Chile
Meghan Cook	Center for Tech in Government, University of Albany, USA
Michael Räckers	WWU Münster – ERCIS, Germany
Michele Osella	Istituto Superiore Mario Boella, Italy
Miguel Garcia Gonzalez	Zabala Innovation Consulting S.A., Spain
Mila Gasco	Center for Technology in Government, University of Albany, USA
Muneo Kaigo	University of Tsukuba, Japan
Nicolau Reinhard	University of São Paulo, Brazil
Nikos Karacapilidis	University of Patras, Greece
Pär-Ola Zander	Aalborg, University, Denmark
Paul Gibson	Telecom Sud Paris, France
Peter Cruickshank	Edinburgh Napier University Scotland, UK
Peter Teufl	A-SIT Plus, Austria
Peter Winstanley	Scottish Government, UK
Radomir Bolgov	Saint Petersburg State University, Russia
Ralf Klischewski	German University in Cairo, Egypt
Ralf Lindner	Fraunhofer ISI, Germany
Renata Araujo	Federal University of the State of Rio de Janeiro, Brazil
Robin Effing	University of Twente, The Netherlands
Roumiana Ilieva	Technical University of Sofia, Bulgaria
Rui Pedro Lourenço	University Coimbra, Portugal
Sabrina Franceschini	Regione Emilia-Romagna, Italy
Saleem Zoughbi	UN, Palestine
Simon Smith	Charles University in Prague, Czech Republic
Soon Chun	City University of New York, USA
Stevan Gostojić	University of Novi Sad, Serbia
Suha Al-Awadhi	Kuwait University, Kuwait

Sylvia Archmann	Counsellor for Governance, Innovation and Professional Formation Training, Austria
Tobias Siebenlist	Heinrich Heine University Düsseldorf, Germany
Todd Davies	Stanford University, USA
Tomasz Janowski	Danube University Krems, Austria and Gdansk University of Technology, Poland
Vanessa Nunes	University of Brasilia, Brazil
Vasily Bunakov	Science and Technology Facilities Council, UK
Wichian Chutimaskul	King Mongkut's University of Technology Thonburi, Thailand
Wilfred Warioba	Bessbrook International Ltd., Tanzania
Witold Staniszkis	Rodan Development, Poland
Wojciech Cellary	Poznan University of Economics and Business, Poland
Yuri Misnikov	ITMO University, Russia

Reviewers

Ann O'Brien	National University of Ireland Galway, Ireland
Arie Purwanto	Delft University of Technology, The Netherlands
Artur Csetényi	Budapest Corvinus University, Hungary
Beatriz Barreto Brasileiro Lanza	PUCPR/CELEPAR, Brazil
Bert Groot	Saxion University of Applied Sciences, The Netherlands
Brianna Smrke	Ontario Public Service, Canada
Carmen Ionela Rotuna	Institutul Naţional de Cercetare-Dezvoltare, Romania
Chris Mason	Swinburne University of Technology, Australia
Christian Voigt	Zentrum für Soziale Innovation, Austria
Deodatus Patrick Shayo	University of Dar es Salaam, Tanzania
Dimitri Gagliardi	The University of Manchester, UK
Fiorenza Lipparini	PlusValue, UK
Francesco Mureddu	Lisbon Council for Economic Competitiveness and Social Renewal, Belgium
Galia Novakova Nedeltcheva	Sofia University, Bulgaria
Gianluigi Viscusi	École Polytechnique Fédérale de Lausanne, Switzerland
Giulio Pasi	European Commission, Spain
Giuseppe Veltri	University of Trento, Italy
Graeme Baxter	Robert Gordon University, UK
Hannes Richter	University of New Orleans, USA
Jérôme Brugger	Bern University of Applied Sciences, Switzerland
Julián Villodre	Universidad Autónoma de Madrid, Spain
Lőrinc Thurnay	Danube University Krems, Austria
Lyudmila Vidiasova	ITMO University, Russia
Mark Melenhorst	Saxion University of Applied Science, The Netherlands

Martina Pasquini	IE University/IE Business School, Spain
Mergen Dyussenov	Lee Kuan Yew School of Public Policy, Singapore
Nettra Pan	École Polytechnique Fédérale de Lausanne, Switzerland
Niels Primus	Saxion University of Applied Sciences, The Netherlands
Peter Winstanley	The Scottish Government, UK
Rodrigo Sandoval-Almazan	Universidad Autonoma del Estado de México, Mexico
Şahika Eroğlu	Hacettepe University, Turkey
Sarah-Kristin Thiel	Aarhus Universitet, Denmark
Sélinde van Engelenburg	Delft University of Technology, The Netherlands
Shefali Virkar	Danube University Krems, Austria
Svein Ølnes	Western Norway Research Institute, Norway
Tarmo Kalvet	Tallinn University of Technology, Estonia
Titiana Ertiö	University of Turku, Finland
Uta Russmann	FHWien der WKW University of Applied Sciences for Management and Communication, Austria
Vitalina Karachay	ITMO University, Russia
Walter Seböck	Danube University Krems, Austria
Yulli Jeremia Bebia	University of Dar es Salaam, Tanzania

Contents

General E-Government and Open Government

General E-Government and Open Government

Suomi.fi – Towards Government 3.0 with a National Service Platform

Jesse Yli-Huumo[1(✉)], Tero Päivärinta[2,3], Juho Rinne[1], and Kari Smolander[1,4]

[1] Department of Computer Science, Aalto University, Helsinki, Finland
{jesse.yli-huumo, juho.a.rinne}@aalto.fi, kari.smolander@lut.fi
[2] Information Systems, Luleå University of Technology, Luleå, Sweden
tero.paivarinta@ltu.se
[3] Department Applied Information Technology, University of Gothenburg, Gothenburg, Sweden
[4] School of Business and Management, Lappeenranta University of Technology, Lappeenranta, Finland

Abstract. The KaPa (Kansallinen Palveluarkkitehtuuri, in Finnish) program establishes the national e-government service platform in Finland. The platform, Suomi.fi, provides a one-stop portal for citizens and organizations to access both public and related private sector services. This research reports a case study of the platform by analyzing it in light of recent characteristics identified with the emerging concept of Government 3.0: openness and transparency, sharing, increased communication and collaboration, government re-organization through integration and interoperability, and use of new technologies. Our results contribute by concretizing the hitherto abstract and loosely defined concept of Government 3.0 by describing a timely and complex national e-government implementation in detail in light of such characteristics. Our study also suggests three emergent themes in relation to contemporary Government 3.0 characteristics: opening up technologies and solutions in addition to open data, cross-border integration and development, and the enhanced role of the private sector in both development activities and merging into the portfolios of one-stop services.

Keywords: E-government · Government 3.0 · Platform · Suomi.fi

1 Introduction

Digitalization of public services continues. In an ideal scenario, the goal of e-government is to create a seamless architecture for public services, where all systems and services are integrated across both the public and the private sectors to provide a one-stop service [1–3] for citizens and organizations. However, while implementing one-stop services have shown to be challenging in municipalities [4] or within particular segments of government alone (e.g. [5]), reaching such a goal at the national level poses even a greater challenge. Indeed, government services may involve several

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 3–14, 2018.
https://doi.org/10.1007/978-3-319-98690-6_1

public sector systems and organizations, which need to be re-structured towards increasing interoperability and collaboration [6].

Emerging technologies will change how governments operate and provide services. E-government development is a continuous process of technical and organizational advancements, which require transformation of services and organizations [7, 8]. Currently, advanced e-government services are often based on use of Web 2.0 technologies, and the contemporary solutions are labeled as Government 2.0 [9, 10]. Recently, the concept of Government 3.0 has been coined to describe the transformation of the public services to address next generation infrastructures, organizational structures, processes and services [11–13]. However, the concept remains still quite unclear, while it has been argued, in general, to promote openness, sharing, increased communication and cooperation in the public sector for citizens, businesses and non-governmental stakeholders [11, 12].

The characteristics of Government 3.0 continue to emerge without clear definitions, which represents a research gap in contemporary e-government research. Meyerhoff Nielsen [13] argues that with such emerging understanding of Government 3.0, it is not yet possible to evaluate whether individual governments are reaching or have reached the 3.0 stage. This calls for more research on examples of recent e-government projects aiming at Government 3.0-related characteristics.

In Finland, which often scores high in international e-government rankings, a majority of the public services are digital, accessible, and widely used by citizens [14]. Finland has a long history with e-government. For example, all branches of the government developed their own systems in the 1970s and the Internet was taken into governmental use already in the 1990s [15]. However, in Finland, alike in many other countries, the implementation of the one-stop service at the national government level has been delayed because of lack of service interoperability and integration.

To address the lack of integration and interoperability and the need to develop a one-stop service for citizens and organizations, the government of Finland started the KaPa (Kansallinen Palveluarkkitehtuuri, in Finnish, National Service Architecture) program in 2014. The main objective was to develop a national architecture for digital services. In this paper, we study the KaPa program and the service platform Suomi.fi developed within the program. Our objective is to understand and explore the impact of the KaPa program to e-government in Finland, and identify characteristics of Government 3.0 in relation to the on-going development. Our research question is "How do the KaPa program and the Suomi.fi service platform promote Government 3.0 in Finland?"

2 Background

Since the mid-2010s, the notion of "Government 3.0" has emerged among policy-makers as a label for next generation ICT-enabled technology innovation and as the successor for "Government 2.0" initiatives [16]. Whereas Government 2.0 refers to the use of Web 2.0 technologies by government [9, 10], Government 3.0 embraces the capture of next generation infrastructure, organizational structures, process and services required for the transformation of the public sector [13]. However, the challenge with

Government 3.0 as a concept is its current loose definition [13]. A recent literature review on e-government maturity models [13] defines Government 3.0 (building on [11, 12]) as: "*Through openness, sharing, increased communication and cooperation the public sector, citizens, businesses and non-governmental stakeholders, the aim is for government to be more service-oriented, competent, and transparent, to proactively provide personalized and customized public services and generate new jobs in a creative manner by opening and sharing government-owned data to the public and encouraging communication and collaboration between government departments*".

The current definitions for Government 3.0 [11–13] seem to be divided into two main perspectives. The first perspective regards use of new technologies as the next generation infrastructure, beyond Web 2.0, to provide better services for citizens and organizations. While an exhaustive list of such emerging technologies remains undone, the mentioned examples include artificial intelligence, semantic web and text analytics, machine learning, internet of things, blockchain, and big data analytics [11–13]. The second perspective is the reorganization of government at several levels. The levels mentioned are infrastructure, organizations, structures, processes and services. Reorganization seems to have two main goals: To increase collaboration and communication within the public sector organizations to remove unnecessary complexity, and to increase openness and transparency of government through concepts such as open data. For a country to reach the "Government 3.0 stage", it is presumably required to take these two aspects in consideration when improving the e-government. Personalized and customized e-government services for the public sector, citizens, businesses and non-governmental stakeholders to gradually reach a one-stop e-government service, should be included in Government 3.0 initiatives. These concepts are currently presented at an abstract level and do not include detailed definitions or descriptions or the particular elements (cf. [13]). However, we drew upon the initial works on Government 3.0 above and formed the following characteristic categories to be used as a conceptual lens on a large-scale national e-government program:

- **Openness and transparency** of government and development
- **Sharing** of data
- **Increased communications and collaborations G2C, G2B, and G2G**
- **Reorganization of government** through integration and interoperability
- **Use of new technologies.**

3 Research Process

Our research is a qualitative, exploratory case study on the KaPa program and Suomi.fi. In our view, KaPa/Suomi.fi represents a rare case [18] of a national, complex program of developing one-stop services for citizens and organizations. We collected data through interviews, survey, observation, and secondary data collection. Firstly, we interviewed 11 KaPa stakeholders. These included KaPa program management, Suomi.fi development team, public sector organizations, and private organizations representing early adopters of the platform. The interviews, which took about one hour each, were recorded and transcribed. Secondly, we performed an online survey in September

2017 with Suomi.fi service platform user organizations (both public sector and private sector). We received 82 responses. The survey was structured regarding experiences with each Suomi.fi service and included questions about general opinions on KaPa as a program. Third, we decided to collect secondary data from news articles and blog posts that were written about KaPa program and Suomi.fi. The number of collected secondary data was +300 items. Last, we took part into two KaPa events to observe stakeholders in the program. The first event was a presentation of online survey results to the Suomi.fi development organization, which also involved two hours of evaluative discussion. The second event was a roadmap event organized by the Suomi.fi development organization to user organizations.

The interviews and events were recorded and transcribed. All data from the interviews, events, survey and news articles and blogs were stored into a database. Two researchers performed data analysis by going through the collected data and creating codes, labels and categories with sticky notes based on collected data. The survey was originally designed for other study, which meant that some of the analyzed data did not have relevance to Government 3.0 concept. After the coding stage and several brainstorming sessions, the identified categories were summarized into higher level themes. While many of the identified themes coincided with the previously (loosely) defined characteristics of Government 3.0 (see Sect. 2), we wanted also to remain open for potentially emerging themes when analyzing the data. The results thus involve a description of five characteristics of Government 3.0 based on the previous literature and three emerging themes (Table 2; the emerging characteristics and themes marked with *).

4 Results

4.1 National Architecture Program in Finland

The Finnish Government initiated the KaPa program in 2014 (budget: 100 million euros) to develop the national architecture for digital services. The main objective was a compatible infrastructure facilitating information transfer between organizations and services. The KaPa program involves a national data exchange layer (based on the x-Road solution used in Estonia [17]); a shared service view required by citizens, companies and authorities; a new model of nation-wide e-identification, and national solutions for the administration of roles and authorizations for organizations and individuals. Suomi.fi is the name of platform and portal where citizens and organizations can access the digital services. The domain (www.suomi.fi) already existed before the KaPa-program, but the previous version served mostly as an information ledger for citizens and organizations to retrieve basic information about public services or forms to fill in. The architecture of Suomi.fi comprises four layers: data, service, interface, and consumer (Fig. 1).

The data layer comprises data registries integrated to the data exchange layer (of the service layer). The data layer is divided further into basic information registries and industry data registries. Basic data registries include key data repositories of the public sector, which are essential for the society to function, such as those of population,

property, companies, communities etc. Industry data registries involve information generated by various industries, both public and private. Integration to the data exchange layer service provides a standardized and secure platform for data exchange among the public and private organizations.

The service layer involves generic services provided through the platform to increase efficiency and innovation in the public sector and in the related private endeavors. The layer involves no end-user services provided by the public and private organizations themselves (for example, tax office service for tax reporting offered to citizens). The service platform provides eight services for organizations (summarized in Table 1), and coordinates obligations and restrictions on service utilization. Service utilization is free.

The interface layer covers all interfaces for organizations and citizens that are used to access the Suomi.fi services. These interfaces can be divided into two types: programming interfaces and user interfaces. The programming interfaces to access the Suomi.fi services are used by both public and private sector organizations. Avoindata.fi and api.suomi.fi are catalogues that present available application programming interfaces (APIs) for integration in Suomi.fi. Service management is for companies to manage Suomi.fi services. User interfaces are used mostly by citizens and organizations to access Suomi.fi. Web services can be accessed by browser or mobile applications.

The consumer layer connects all users to the Suomi.fi services. Users are divided into two main categories: service providers and ecosystems, and citizens. Service providers and ecosystems include the public sector and private organizations using Suomi.fi. These organizations can be Finnish or from the European Union area. Citizens involve both Finns and other EU citizens, who wish to use the services in Finland.

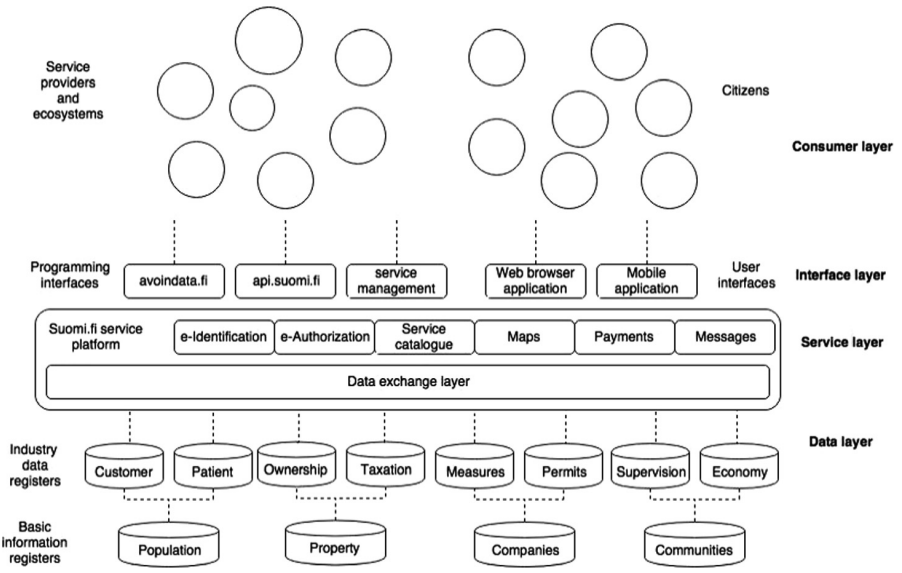

Fig. 1. Suomi.fi architecture

Table 1. Suomi.fi platform services

Service description	Availability
e-Identification enables organizations to authenticate their service users with strong electronic authentications methods; A single sign-on for citizens, which provides access to all public sector services (that use e-Identification)	Public sector organizations are eligible to use the service (organizations that require strong authentication are obligated by law). In principle, private sector organization do not have a right to use e-Identification service, unless they provide services to public sector
e-Authorization enables citizens or organizations to authorize another citizen or organization to act behalf of them	Both public and private sector organizations are allowed to use e-Authorization
The data exchange layer enables standardized and secured data exchange between organizations (based on x-Road [17])	Both public sector and private sector organizations are allowed use and provide data (Public organizations with public data registers are obligated)
The service catalogue enables organizations to describe their services in a standard way to a common database	Both public and private sector organization are allowed to use service. Organizations that produce public services are obligated
Maps service enables a centralized way for organizations and citizens to view and present locations	Public organizations are allowed to use maps. Private organization are not allowed to use the maps service, unless they provide services to the public sector
Payments enable public organizations to send invoices to citizens. Citizens can access payments through Web portal	"Vital" -organizations, such as tax office, are obligated to use the service. Other public sector organizations are allowed to use the service. Private sector organization are not allowed to use this service, unless they provide services to public sector
Messages are operated to serve citizens and organizations alike. Citizens can access messages through Web portal	Public sector organizations are either obliged or allowed to use the service (cf. payments). Private sector organizations are not allowed to use the service, unless they provide services to public sector
Web portal is an integrated view on public services (combining all Suomi.fi services under one view). Organizations can provide registers of their web services that citizens can view through one portal	Public sector organizations are either obliged or at least allowed to provide registers. Private sector organization are allowed to use this service, if they have rights to handle citizens' social security numbers or business IDs

4.2 KaPa Program in Light of Government 3.0

Table 2 summarizes our analysis of the Kapa program under the themes of the previously described Government 3.0 characteristics (Sect. 2) and three new emerging themes (marked with *). One of the new themes, openness of source code, is discussed an already existing Government 3.0 characteristic (openness and transparency).

Table 2. Government 3.0 themes in KaPa

Openness and transparency
Public sector organizations databases and registries opened for the public; ***Source code available at GitHub for anyone to freely review, copy and use***; Suomi.fi development process is transparent to the public through continuous updates with news and blogs from the development team; API catalogues available on the Internet for the public
Sharing
Government-to-government and government-to-business organizations have a possibility integrate their systems to provide, acquire and exchange data with other systems through the data exchange layer
Increased communication and collaboration
Government-to-citizens: Before KaPa, Suomi.fi was only an information ledger and citizens needed to access governmental services through a plethora of separate portals and web sites. The new Suomi.fi is a personalized and customizable one-stop service for citizens with a single sign-on to public services. **Government-to-businesses:** Before KaPa, Suomi.fi was only an information ledger and included some forms and documents needed to start or run a business. New Suomi.fi provides opportunities for businesses on leveraging free public administration databases and services to innovate new services. **Government-to-government:** Before KaPa, the public sector organizations operated in several separate networks. New Suomi.fi provides data sharing and interoperability of information systems across the public administration
Reorganization of government
Before KaPa, e-government consisted of independent components and infrastructure was scattered. With KaPa, the production model turns from vertical to horizontal. Service providers do not need to produce all the layers below the service (such as servers, service capacity, databases and integration solutions), so the service provider can concentrate on the top layer (application). Purpose of KaPa is to support the national economy by making public administration more efficient and by creating new business opportunities in the private sector
Use of new technologies
Use of new technologies (those mentioned in the Government 3.0 literature) was not evident in the KaPa program. Even though we identified some mentions in our data related to e.g. use of blockchain or data analytics in e-government services, it was not seen necessarily the main objective of the KaPa program but more as a task of the future
**Cross-border government-to-government integration*
The KaPa program involves also an international government-to-government platform integration with Estonia. The integration between the two platforms represents, to our knowledge, the first nation-to-nation integration at this level and a step towards Pan-European e-services
**Increased private development innovations on public services*
The private sector is also involved in Suomi.fi service platform development; The role of the private sector is regarded as important in new services development and innovation through the Suomi.fi platform

Openness and Transparency. The concept of openness and transparency was mentioned in several data sources. First, the Ministry of Finance of Finland, the responsible entity of KaPa, declared in their program statement that one of the main objectives of KaPa is to "*advance the concept of openness in public sector*". In practice, this meant

that public databases and registries were opened and made available through the data exchange layer to other public sector organizations and private sector businesses. As an example, it is possible to retrieve data about vehicles from Finnish Transport Safety Agency or from a public Finnish trade register that contains information on traders and businesses from Finnish Patent and Registration Office. Currently, the data exchange layer does not yet include all public sector databases and registries. Second, Suomi.fi development was promoted as open source development, which led to making source code of services available freely on GitHub. Anyone can see the source code of services such as x-Road (data exchange layer) and the e-Identification service. Third, the development team provided frequent news and blog posts about the status and development of Suomi.fi concerning the program advancements, schedules, deadlines and challenges, through an information channel called esuomi.fi. Fourth, APIs of Suomi.fi that are available for service developers are listed on public API catalogues in Internet.

Sharing. One of the main goals was to use the Suomi.fi platform as a data exchange layer among the public and private sector organizations alike, which represents an important theme of Government 3.0. All organizations are entitled to join the data exchange layer to use and provide data from registries and databases. According to a development organization representative, the data exchange layer helps retrieve data by applications that need to integrate several data sources. Data exchange enhances efficiency of public services through boosted interoperability. Private businesses can also create new services through open data from the public databases and registries. In the interviews with Suomi.fi user organizations, the data exchange layer was described as "*a vital element for interoperability, because otherwise systems wouldn't communicate between each other*", or as an interviewee from a private company expressed, "*we need data from the basic governmental registries*". However, doubts and critique were also expressed. An online survey on the user organizations revealed a few comments such as: "*At the moment there are no incentives to join the platform*". Many organizations did not yet recognize the business needs and benefits available on the current data exchange platform. One of the reasons might be the current number of organizations joined to the platform (on 22th Feb 2018, the number public and private organizations combined was 81). For example in the Estonian x-Road, the number of active organizations is around 1000 [19]. On the other hand, Suomi.fi is still in an early stage of implementation, and the number of organizations is expected to increase.

Increased Communications and Collaborations in G2C, G2B, and G2G. A representative of the Ministry of Finance described Suomi.fi as a "*Shopping mall for citizens to access public services*". The Finnish Government provides Suomi.fi platform and technology (shopping mall), the service providers (both public and private) use Suomi.fi platform to provide the services (shops), and citizens and organizations enter Suomi.fi platform to use provided services (customers). Compared to the previous Suomi.fi, which only served as an information ledger to citizens and businesses, the new Suomi.fi provides increased communication and collaboration between citizens, businesses, and government. The model and platform is expected to implement a customizable and personalizable one-stop service view for citizens and organizations. The new platform is argued to remove the problem where "*in every e-government*

service you need to log in separately, and it is jumping constantly from one place to another" (Manager, Ministry of Finance). The objective of KaPa is to create gradually a service where all public sector services and many private sector services can be accessed by citizens to create a customized and personalized view. "*At the end, citizens select, what set of services they want to collect to their service view*" (Manager, KaPa)

Reorganization of Government Through Integration and Interoperability. The concept of increased integration and interoperability between government, public sector and private sector systems was a visible theme also in KaPa. "*KaPa program fixed the issue of lack of IT-systems integration. One of the most important features of the KaPa program is the centralized data exchange layer, which means that systems can communicate with each other. This will improve interoperability of public sector systems tremendously*" (manager, KaPa). Instead of individual point-to-point integrations between systems, the new platform strategy will enhance many-to-many system interoperability. Besides the data exchange layer, also new e-Identification service will increase interoperability among the public sector services. The new Suomi.fi platform enables single sign-on to the services. However, coordination among the service providers and the platform involves also challenges. "*One of the challenges of taking service in use is that organizations think that Suomi.fi services are good overall for the public sector, but do not see the need for their own operations*" (Manager, KaPa). Therefore, the platform coordinator needs to introduce good support practices to manage change and to gather on-going feedback from the user organizations.

Cross-Border Government-to-Government Integration. Finland and Estonia use the same technology (x-Road) for the data exchange layer. The x-Road technology was donated to Finland by Estonia for free and adopted as a basis for Finnish developments. "*We are using the software which was originally from Estonia X-road, and have done quite a lot of development on top of it*." (Manager, Suomi.fi development). A jointly established institute by Finland and Estonia develops the digital data exchange layer of x-Road technology further. In addition to co-operation in development, the two countries integrate also their very services and data exchange through the joint platform across borders [19]. "*Integration makes it possible to exchange data and to use services between two data exchange layers, as fluently as done within Suomi.fi alone*" (interviewee, Ministry of Finance). The cooperation between Finland and Estonia can be regarded as a step towards the Pan-European e-services, introducing an interesting aspect of e-government development in the both countries.

Increased Private Development Innovations on Public Services. E-government is, by definition, mostly associated with the public sector organizations. In this case, however, the role of private sector businesses in the KaPa program and development of the Suomi.fi platform was brought into attention many times. The KaPa program goals stated one of the goals as "making public sector more efficient and creating new business opportunities for the private sector." This notes the current importance of private companies in Finnish e-government. Even though a few services were restricted from the private actors due to legal constraints, such as e-identification, maps, payments, or messages, the data exchange layer, e-authorization and service catalogue represent new opportunities of service innovation to private businesses. These services

are free to use, which is expected to attract the private actors as well. The private sector is seen also as a necessary success factor for the platform, "this will only be a good project if private companies decide to join" (Project Leader, Suomi.fi). Moreover, development of Suomi.fi took place in collaboration between public organizations and private companies. The KaPa program management was coordinated by the government departments, but the actual development of Suomi.fi platform was mostly conducted by 3rd party development teams. Based on these observations, we regard the role of private sector businesses in e-government development and integrated service delivery as an emerging theme to complement the recent concept of Government 3.0.

5 Discussion

Our results suggest two main contributions to the previous literature on the emerging concept of Government 3.0. Firstly, the case analysis showed that the KaPa program and Suomi.fi involve the most conceptual characteristics identified with the recent concept of Government 3.0 (cf. [13]). Hence, we regard Suomi.fi as a rare case which as such illustrates how many of the (hitherto loosely defined) aspects of Government 3.0 can be concretized within a nation-level program and a platform aiming at one-stop public services for citizens and businesses. Especially, our analysis of Suomi.fi illustrates how it concretizes the concepts of openness and transparency, sharing, increased communication and collaboration at different levels of governance interactions (G2C, G2B, G2G), and re-organization of government.

Secondly, our analysis suggests three emergent themes that complement the contemporary characterizations of Government 3.0 (if compared to [13]): 1. Opening technological components and platforms (not just data) of public-sector solutions for further utilization and innovation, 2. cross-border cooperation on service development and integration, and 3. Increasing integration of the private sector in service innovation and delivery; both in development processes and innovations of the platform as such and as introducing a potential solution for private service providers to participate in the service delivery portfolio.

Suomi.fi, together with Estonian x-road, provides an arena for open APIs that enable distributed development and increased contributions from the private sector to service portfolios, and even to the technological infrastructure. In addition to increased data accessibility and open data initiatives this should create a plethora of new innovations on Finnish e-government in the near future. The new level of integration of cross-border collaboration between Finland and Estonia will produce both more efficiency and create also possibilities for new innovative G2G interactions. Moreover, the private involvement was seen as a crucial success factor for future success of the platform. While we argue that our research contributes by identifying these three themes and discussing about them in the Finnish context, there exists a good number of new research avenues on both organizational and technological advancements in government, as well as the socio-political impacts of such developments over time. Such developments should be analyzed simultaneously from several viewpoints of political, organizational, and technological opportunities and challenges, which will most likely only start to emerge.

The current Suomi.fi concept includes few or no signs of "new technologies" as envisioned in the contemporary ideas of Government 3.0 [12, 13]. Hence, we do not yet claim that Suomi.fi would fully represent Government 3.0 in the Finnish context. However, we see the platform and portal development rather as an enabling step towards such developments and predict that the technologies mentioned in the contemporary Government 3.0 literature will be increasingly adopted, and enhanced by the integration platforms such as Suomi.fi, if taken into use. For example, data integrations across government agencies provide new opportunities to apply artificial intelligence, semantic web and text analytics, machine learning and big data analytics -related technologies for innovating new services and knowledge building on the public data. IoT-based services and subsequent data exchange integrate ever better to the standardized service platforms, and a whole cluster of new innovative stakeholders can take part on dynamic service development with openly available new technologies and data.

According to our analysis, Finland has thus taken steps with the KaPa program towards Government 3.0. However, there are also other leading-edge national e-government countries such as Estonia [17], Norway [20], Netherlands [20], South Korea [9, 12] and many others. To address future research avenues, more in-depth cross-border comparisons and validations for a stage model for Government 3.0 are required. We believe that the future research needs to focus on following the adoption and impact of new technologies in connection to integrated and interoperable national e-government solutions. Time will tell how control and coordination mechanisms for both increased cross-border (e.g., Pan-European) cooperation and seemingly increasing private sector involvement blurring the borderlines between the public and the private in some countries become implemented. In this regard, a thorough political debate on the eventual political consequences of the increased service and information processing integration across the national as well as public-private borderlines needs to continue.

6 Conclusions

Government 3.0 has been coined to describe the next generation transformation of e-government. The case study on Suomi.fi illustrated how many of the hitherto loosely defined characteristics of the Government 3.0 concept were concretized in the Finnish context. In addition, the results suggests three emergent themes to be scrutinized in Government 3.0 initiatives: opening up technologies and solutions in addition to open data, cross-border integration and development, and the enhanced role of the private sector in both development activities and merging into the portfolios of one-stop services. However, we did not identify signs of new technologies, such as AI, machine learning and data science, as envisioned in the contemporary concept of Government 3.0. We argue that an integrated service platform, such as Suomi.fi, could be seen as an enabler for adoption of new technologies and thus a necessary step on the way towards Government 3.0. For future research, we call for more in-depth comparisons of cases and validations for a stage model for Government 3.0 and analyses of the impact and success of new technologies in connection to integrated and interoperable national (and eventually, international) e-government solutions.

References

1. Layne, K., Lee, J.: Developing fully functional E-government: a four stage model. Gov. Inf. Q. **18**(2), 122–136 (2001)
2. Tat-Kei Ho, A.: Reinventing local governments and the E-Government initiative. Public Adm. Rev. **62**(4), 434–444 (2002)
3. Gouscos, D., Kalikakis, M., Legal, M., Papadopoulou, S.: A general model of performance and quality for one-stop e-Government service offerings. Gov. Inf. Q. **24**(4), 860–885 (2007)
4. Signore, O., Franco, C., Maurizio, P.: E-government: challenges and opportunities. In: Proceedings of the CMG Italy XIX Annual Conference, pp. 1–16 (2005)
5. Schuppan, T., Koehl, S.: One stop government: stalled vision or a matter of design? – empirical findings from social services in Germany (2017)
6. Yannis, C.: Interoperability in Digital Public Services and Administration: Bridging E-Government and E-Business. IGI Global, Hershey (2010)
7. Andersen, K.V., Henriksen, H.Z.: E-government maturity models: extension of the layne and lee model. Gov. Inf. Q. **23**(2), 236–248 (2006)
8. Lee, J.: 10year retrospect on stage models of e-Government: a qualitative meta-synthesis. Gov. Inf. Q. **27**(3), 220–230 (2010)
9. Nam, T.: Government 3.0 in Korea: fad or fashion?. In: Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance, New York, NY, USA, pp. 46–55 (2013)
10. Sun, P.-L., Ku, C.-Y., Shih, D.-H.: An implementation framework for E-Government 2.0. Telemat. Inform. **32**(3), 504–520 (2015)
11. Charalabidis, Y.: What is government 3.0?. In: Charalabidis, Y. (ed.) Governance and Transformation. Yannis Charalabidis, Athens (2015)
12. Ministry of Interior Korea: Government 3.0. Ministry of Interior Korea, Seoul (2016)
13. Meyerhoff Nielsen, M.: Governance failure in light of Government 3.0: foundations for building next generation eGovernment maturity models. In: Ojo, A., Millard, J. (eds.) Government 30 - Next Generation Government Technology Infrastructure and Services. PAIT, vol. 32, pp. 63–109. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63743-3_4
14. European Commission: eGovernment in Finland, February 2016, Ed. 18.0 (2016)
15. Verohallinto: Verotuksen historiaa Suomessa. Verohallinnon Julk. 38109 (2009)
16. Hassan, I.A., Adegboyega, O.: Enabling gov 3.0 through semantic web, natural language processing and text analytics. In: T-Gov Workshop 2014, 12–13 June 2014, vol. 14, pp. 64–81. Brunel University, London (2014)
17. Anthes, G.: Estonia: a model for e-Government. Commun. ACM **58**(6), 18–20 (2015)
18. Yin, R.K.: Case Study Research: Design and Methods. Sage Publications, Thousand Oaks (1989)
19. eSuomi.fi. https://esuomi.fi/suomen-ja-viron-palveluvaylat-liitetty-yhteen-7-2-tietojen-nopea-ja-luotettava-vaihto-maiden-valilla-nyt-mahdollista. Accessed 08 Feb 2018
20. Aagesen, G., van Veenstra, A.F., Janssen, M., Krogstie, J.: The entanglement of enterprise architecture and IT-governance: the cases of Norway and the Netherlands. In: 2011 44th Hawaii International Conference on System Sciences, pp. 1–10 (2011)

Understanding an Integrated Management System in a Government Agency – Focusing Institutional Carriers

Ulf Melin(✉), Karin Axelsson, and Truls Löfstedt

Department of Management and Engineering, Information Systems, Linköping University, 581 83 Linköping, Sweden
{ulf.melin, karin.axelsson, truls.lofstedt}@liu.se

Abstract. Working with an integrated management system (IMS) is a challenging task. In public organizations, the formalization of an IMS including the communication of control mechanisms, rules, goals and culture are crucial. Several types of carriers are used in order to communicate the content in an IMS – both human actors and artefacts. An artefact studied in this paper is an intranet, as one carrier of the IMS. The purpose of this paper is to explore how institutional theory – focusing institutional carriers – can help us to understand how an IMS is represented through human actors and technology in a government agency. The conclusion is that the application of an institutional carrier perspective on an IMS can help us to understand the past and present, the role, and the relative success of such a system. An IMS can be aligned or misaligned related to three dimensions of structure, process and people. Achieving an aligned and legitimate IMS is crucial in order to achieve goals in an organization. The implications of this study are that further research and practice should give more attention to institutional carriers when studying and improving IMS.

Keywords: Integrated management systems · Governance
Institutional theory · Institutional carriers · Government

1 Introduction

Management systems of different kinds are used in many contemporary organizations. In government organizations of today in an era of New Public Management it is commonplace. There are general management systems used in order to align individuals' activities with the overall organizational goals and mission [22] in an act of organizational coordination [23, 24]. A general definition of a management system is that it consists of "[…] a set of procedures an organization needs to follow in order to meet its objectives" [17]. There are also dedicated management systems for management control, IT governance, quality, environment, etc. Malmi and Brown [22] define management control systems as packages, which means "[…] a collection or set of controls and control systems." (ibid., p. 287). Such systems can be linked to traditional accounting controls (e.g. budgets), administrative controls, and more socially oriented controls (e.g. values and culture) [22]. In this paper, we study a management system that is packaged and spans over a whole government agency. It contains accounting,

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 15–28, 2018.
https://doi.org/10.1007/978-3-319-98690-6_2

administrative, and socially oriented control sub-systems. Since the system is considered as a package, spanning over the organization as a whole, with this integrative aim, we define the management system as *integrated*. Working with an integrated management system (IMS) is a challenging task in theory and in practice [18]. The complexity in the management system pose a challenge, so does the representation, use and development of it. In any organization, the formalization of the IMS and the communication of control and coordination mechanisms, rules and regulations, goals, culture and values are crucial in order to achieve goals. In the public sector we also need to recognize the complexity of values in an IMS, with e.g. competing values in internal efficiency and citizen benefit [2, 4, 31].

In this paper, we explore the notion that several types of institutional *carriers* [33, 34], linked to pillars of institutions, can be used in order to analyze an IMS. This is done through the study of influential stakeholders (the management group) and the use of an intranet supporting the IMS in a government agency focused as a case in this paper. The use of carriers is a concept that stems from an institutional theory perspective. Based on an institutional perspective, information systems (IS) are artefacts carrying various types of institutionalization [16, 33]. An IMS can also be carried by human actors, as individuals or in groups. Studying IT artefacts from this perspective within the IS domain is common and analytically promising [11, 12, 16, 28, 30], but fewer studies have explicitly explored the expressions of particular IT artefacts, like the intranet as one carrier of the IMS in our case. An institutional perspective can help us to understand the processes promoting and impeding change in organizations; processes informed by historically guided values and norms that have stabilized and persisted [29, 30]. One of few studies exploring IT artefacts, focusing on an intranet, from this perspective is provided by [5] discussing the paradox between institutionalization and awareness of the strategic value of the intranet in an organization. The authors highlight what happens when IT is embedded in the organization, as is the case with an intranet, by using an institutional theory perspective. However, they do not explicitly discuss this from an IMS perspective. To study the use and the development of an IMS from an institutional perspective is therefore interesting since it potentially can help us to understand the role of values, norms and the use of different carriers to realize the intentions of a management system when coordinating a government agency. Doing this, path dependency connects the past and the future by understanding the role of institutionalism [34, 41].

Thus, the purpose of this paper is to investigate, show and reflect upon how institutional theory – focusing on institutional carriers – can help us to understand how an IMS is represented through human actors and technology in a government agency. By applying the institutional theory perspective to the studied IMS implemented and carried by the agency's management group and an intranet we aim to understand the past and present, the role, and the challenges and relative success of an IMS. The research questions, asked, are the following: (1) how can we understand an IMS, carried by actors and IT artefacts, and its role in coordinating a government agency, and (2) what can we learn from applying institutional theory, focusing on carriers, pillars, and path dependencies from this case? The expected contributions from this paper are to broaden the scope on how to understand management systems applying an

institutional perspective, to reflect upon the challenges and relative success of an IMS from this point of departure and the implications for future research and practice.

After this introduction, we introduce institutional theory and carriers, together with IMSs and governance in Sect. 2. In Sect. 3, we define and discuss the research approach followed by a case study introduction. Then we analyze the case by applying the institutional theory perspective to the empirical findings and formulate lessons learned from the case study in Sect. 4 which we conclude in Sect. 5.

2 Previous Research

Below we introduce previous research on institutional theory and technology, together with integrated management systems and governance, that we also use to frame and to analyze the IMS in the case study.

2.1 Institutional Theory and Information Technology

Institutional theory [29, 32] have proven to be powerful when understanding and conceptualizing the complex nature of technology, institutional arrangements and its embeddedness in socio-economic contexts [14]. Institutions can be seen as structural arrangements that guide and restrict actors' behaviour (Berger and Luckmann [7]). Rules of behaviour is an important dimension and range from cultural and mental models to rules in terms of legislation as well as from norms to political structures [34]. When studying and understanding institutions *path dependency* is an important concept connecting the past and the future in this line of research [34, 41]. Path dependency can be described as increasingly constrained processes that cannot easily be escaped when changing organizations. The concept is also considered as one answer to the question of how and why *institutionalization* occur [34] – institutionalization in terms of externalization, objectification and internalization [7].

Robey and Boudreau [30] specifically suggest the use of institutional theory to understand the processes promoting and impeding change in organizations. IS researchers trying to explore the complex interplay between IT and its social context have also promoted and used institutional theory as a vehicle [6, 11, 16, 28]. If we take a closer look at the digitalization of government, the role of institutional theory as a perspective to understand the complex interplay in this sector and research domain is acknowledged by e.g. [14, 21]. From previous research, we also know that government organizations are operating based on a set of public values affecting what is possible and suitable to achieve in general and linked to the use of IT (see e.g. [4] or [31]). We choose to explore the empirical data in this paper using institutional theory, focusing pillars, carriers and path dependency, since it sheds light on the IMS and how it is carried by actors, structures and artefacts in the organization and how this has evolved over time. Based on this perspective, information systems are artefacts carrying various types of institutionalization [16, 33]. Thus, these artefacts are filled with assumptions about how government agencies should be organized. IT artefacts are not neutral; they contain additional ideals ascribed to them. As non-neutral artefacts they can support or impede the original intentions.

2.2 Institutional Pillars and Carriers

Three elements – pillars – have been considered as vital ingredients for all institutions [34]; (1) regulative, (2) normative, and (3) cultural-cognitive systems. For a more detailed description of these elements we refer to Scott [34]. Elements in institutions are conveyed by different types of carriers expressed as follows: "Institutional ideas travel in many ways. Not only are they made up of various elements […] but they are carried by several types of vehicles." [33, p. 882]. Carriers are presented in different versions and with different labels. By using Scott [34] as a point of departure we can identify the following classes of carriers: symbolic systems, relational systems, activities, and artefacts. Institutional carriers are linked to the regulative, normative and cultural-cognitive pillars [34]. Symbolic systems can, for example, carry regulative elements (rules and laws), normative ditto (values, expectations, and standards) and cultural-cognitive elements (categories, typifications, schemas). Relational systems can carry regulative elements (governance and power systems), normative elements (authority systems), and cultural-cognitive ditto (structural isomorphism, identities). Activities as the third type of carries can be a vehicle for regulative elements (when e.g. monitoring and sanctioning), normative elements (like roles, jobs, routines and habits), and cultural-cognitive elements as a pillar like predispositions and scripts. Finally, artefacts as the last carrier can serve as a vehicle for regulative elements (objects in line with specifications), normative (objects meeting standards), and cultural cognitive elements (objects with symbolic value) [34, p. 95 f.]. Scott [33] also states that carriers are interdependent and can be combined. In the analysis below we use the different types of carriers as a theoretical perspective to investigate how an IMS is carried by a management group and an intranet. One important statement made by [33] is that the carriers as vehicles are not neutral; they effect the elements that are transported. Viewing IT artefacts as carriers is put forward by [6, 27].

2.3 Integrated Management Systems and Governance

The broad definition of management systems provided by ISO [17] can be elaborated in several different categories of management systems. Organizations also develop specific management systems over time, gathering their own processes and routines for managing certain operations. ISO notes that some systems are formalized while other systems exist entirely in the minds of individuals in the organization. Organizations that use several management systems often seek to integrate them [35]. The collected description of all operations inside an organization can be described as several connected management systems or an IMS. The idea of integrating management systems can be traced back to the desire to create management structures that are congruent and easy to maintain [8, 15]. Another motive is to enable the best use of shared resources to meet an objective. This is another challenge that is clearly linked to the case studied in this paper – if and how to integrate several management systems into one system (an IMS).

Jørgensen et al. [18] elaborate on the work of Wilkinson and Dale [43] and suggest a three-level model where an organization can seek to integrate management systems; "(A) Corresponding: increased compatibility with cross-references between parallel

systems. (B) Coordinated and coherent: generic processes with focus on tasks in the management cycle. (C) Strategic and inherent: an organizational culture of learning, continuous improvements of performance and stakeholder involvement related to internal and external challenges." [18, p. 714].

Management systems and the need to strategically govern processes and businesses using different IT artefacts are important in contemporary organizations in different sectors. In order to do this successfully and in a balanced way strategic alignment is suggested as one approach. Early studies emphasize that a higher degree of alignment between business and IT strategies will result in more successful organizations [40]. Strategic alignment and its results have been well-researched over the years, as broad literature reviews [3, 10, 38] show. In recent studies the notion of IT governance has been related to strategic alignment issues. De Haes and Van Grembergen [13], for example, illustrate how IT governance has been implemented in many organizations in order to improve the alignment between business and IT. Management systems can be a tool to coordinate and control these activities and components. An organization's structures, processes, and people are important carriers of the management system, which will be further discussed below.

3 Research Approach and Case Study Introduction

This single, qualitative and interpretative case study [cf. 25, 42] focuses on developing empirically and theoretically grounded knowledge about IMSs in a government agency setting. This paper is written in the context of a project aiming at studying and evaluating the current management system and contributing to the foundation of a new IMS from a research-based perspective. We categorize this study as an act of engaged scholarship [39] where we combine theorizing and contributions to the government agency's problem solving.

Qualitative data collection and analysis was conducted in an iterative and reflexive manner in three stages starting with document studies to get an overview of the current management system (1). Twelve semi-structured interviews [26] where then (2) conducted over a period of two months (Jan–Feb 2015) with representatives from several business areas and hierarchical levels at the headquarters of the organization (Finance and control, Safety management, Performance management, Controlling, Information management, Communications, Operations). The semi-structured interviews were held in Swedish and quotes used in this paper are translated to English by the authors. The interviews were guided by open ended questions. In some of the interviews respondents used a live version of the intranet to demonstrate some of the IMS features and how they interpreted the implemented documents and processes. That gave us the opportunity to observe the intranet in use and how the IMS was implemented in terms of e.g. representation, design, document structure, search engines, etc. This part of the data collection was documented with a camera. Alongside this process (3) we conducted a hermeneutic literature review [9] to increase our understanding of the current situation and explore themes that surfaced continuously during the analysis of the data. The analysis was performed during the research period where interviews were transcribed and the responses were categorized inductively, as a part of a content analysis

approach (cf. [20]). The main categories identified in the analysis were; perceptions of an IMS, structures, roles and responsibilities connected to the IMS, and the role of the IMS in the organization (how the IMS is intended to be used and is used throughout the organization). This is an example of a reflexive research process [1] working with the generation of categories based on the empirical data while using theory as a guide (e.g. institutional carriers) when analyzing data [42]. In this paper, we focus on the latter part in this reflexive research process using institutional theory as a lens and the themes emerging inductively from this analysis. In March 2015, we also conducted a workshop to validate the collected data, the initial analysis and also gathered additional data. This workshop was attended by seven representatives from the organization. A final workshop was held in June 2015 in order to present and discuss results as a part of the engaged scholarship approach mentioned above.

The setting for this case study is a government agency in Sweden with over 5000 employees spanning over several operational areas. The annual budget is over 5.5 Billion EUR (2014). The large and heterogeneous government agency uses an IMS that fills several functions in the organization. According to our respondents the IMS should depict the organizations structure and processes (and contact points in and in-between). The intention from the management group with the IMS is that it also should communicate "soft" aspects such as culture and norms. As such, the IMS should work as both control and guidance throughout the organization and in some situations also in collaboration with external resources. Influential stakeholders, such as the management group is one important carrier of the IMS, so is the intranet – the window of the IMS also carrying it and making the IMS accessible throughout the organization.

Currently the IMS consists of documentation of all operations in the organization. All the documents (approx. 4000) are published and accessible through the intranet where an overview of each business function is provided for employees. In most parts of the organization, the intranet has come to represent (become the window of) the management system as a whole. Specific actors and roles related to each business unit are responsible for keeping the documents related to their function relevant and up to date. The process of doing this is, however, carried out differently across functions. This has led to inconsistencies in notation, detail and depth of the descriptions. There are also differences in the perception of the IMS. The top management perceive the system more as a control system, whereas the employees view it as a guiding tool. This is one of the challenges that guided the analysis below.

4 Analysis

In this section we present our analysis of the IMS, based on institutional theory. The analysis is organized in two themes emerging from this paper's focus on how the IMS is carried: (1) the IMS and the management group carrying the IMS, and (2) the intranet carrying the IMS. The findings are summarized together with lessons learned from the case study linked to previous studies covering e.g. the integration dimension of the IMS, alignment, and the institutional perspective with a key focus on carriers, pillars and path dependency.

4.1 The IMS and the Management Group Carrying the IMS

The managerial attention toward the IMS emphasizes the necessity to have an IMS. The IMS can be characterized according to a general ISO definition [17] and can be considered as a package of controls following [22]. However, the top management and several management levels at the government agency clearly state that it is important that there is *one* integrated IMS present at the agency – not a package of different management systems. The IMS can have different controls, but it should be considered as one system. The reason for putting a large emphasis on the necessity of having only one IMS can be understood as a clear expression of *path dependency* [34, 41]. The agency merged from two rather different agencies historically (with different structures, process and organization culture) and there has been important to govern the agency as *one* organizational unit since then – not two. This explains the expressed need to have only *one integrated IMS* according to the management group, carried also by one united management group. However, the intention to have one IMS in order to merge and coordinate the agency is more alive in the strategies and the rhetoric surrounding it, than in practice. In order to realize this intention, the management group needs to carry the IMS more explicitly. This can be linked to the symbolic, relational and activity dimensions of institutional carriers. Symbolic dimension can e.g. be used to realize the regulative elements from the institutional pillars containing rules in the organization covering the need to use the IMS, normative dimensions containing values, expectations and standards (below) within the IMS communicated from management to co-workers and cultural-cognitive elements such as typifications within the IMS for different usage situations, etc. Relational systems linked to the management group carrying the IMS is a central part of the overall governance structure and processes within the government agency. Establishing legitimacy for the IMS through management's own activities following the prescribed processes (in line with e.g. prescriptions within the IMS) and rules for example is important.

The intentions linked to the IMS are multi-faceted. Listing the intentions explored in the interview data and from the document studies, empirical data involves aspects from supporting employees in their daily work, via enablers for control and monitoring, safety, legal compliance, communicating overall strategic objectives and visions, providing information on news, supporting standardized process, to describing the organization as a whole and as a vehicle to enable an active leadership. Contrasting these intentions with other interpretations and voices elaborating on whether the top management at the agency actively carries the IMS, we have identified that it is considered that the management's actions and attention are not always in line with the guidelines and the set of objectives for e.g. legitimate behaviour in the IMS. This means that the expected dimensions of carriers introduced above regarding the symbolic, relational and activity parts are not efficiently or explicitly present in management behaviour of today in the case. Being a legitimate carrier of a "successful" IMS requires compliance and a synchronized picture in these dimensions. One respondent at the agency expressed this in the following way when asked what the main objective of the organization was:

> *"Our main objective? […] is that what is written on paper [the IMS], or what is communicated by the management?" (Business Developer, February 10, 2015)*

In this example, we can see that adherence to the IMS is related to how *the management group carry the IMS*, how they act, and to the characteristics also of the intranet carrying (cf. [34]), communicating and legitimizing it. Regulative, normative and cultural cognitive elements of the IMS are clearly affected by different carriers, in this case human actors in different management roles in a relational dimension of the system. This was discussed during a workshop arranged at the agency. This was considered as a discrepancy between the content in the IMS and the image of the IMS and the management staff, on different levels from top to middle management. A common view here is that the management needs to have a more unified view of the IMS in the organization, the legitimacy of it, and the representation of the same system linked to different carriers (above). There is also a line of thinking describing that people in middle management and more operative roles needs to develop a further understanding of the IMS, and also legitimize it in thoughts and in daily operative work; to use the IMS as an aid to reach the overall goals in the organization. For example, people in the organization are all parts of a relational system where normative and cultural-cognitive (shared values) elements are formed and traded over time [34].

Managers carrying an IMS through behaviour and an active use of the intranet carrying it, could create a more efficient and solid alliance between human actors, behaviour and technology *strengthening* and legitimizing the IMS, instead of *weakening* it as illustrated above. This can be expressed in terms of using the directive power of an IS as an institution, and be traced back to cognitive and normative elements that are embedded in such artefacts (cf. [16]). Rules with *normative forces* – a kind of prescriptive behaviour – are also embedded in the system directed towards designated positions in the organization and its actors (ibid.). Starting with the management positions as examples of important roles, illustrating prescriptive behaviour and its embeddedness in the intranet carried IMS, can make the IMS more active and efficient in the organization.

4.2 The Intranet Carrying the IMS

In the case study, it was evident that the intranet was an important *carrier* [34] of the IMS. The intranet was the major representation of the IMS for many people in different roles in the heterogeneous organization performing their work. To some extent the artefact was even synonymous with the IMS (cf. [6, 27]) – the *media* (the artefact) and the *content* focusing the different parts of the IMS were treated as one by many employees and users in the organization. When several of our respondents in the interviews and in the workshops described the IMS, they even described the functional and graphical user interface design of the intranet and thereby put the artefact in the foreground rather than the IMS. They also pointed out the importance of the design of the intranet as an *artefact* and a *symbolic* system [34], and how important it was in how well key information (regarding e.g. the support for routines in their daily work) was communicated in the organization as a part of management and control. This puts a lot of pressure on the overall design of the intranet being effectively designed in order to carry the IMS appropriately. Respondents kept coming back to the challenges of the design and underlying logic involved in searching for information in the IMS carried

by the intranet, and to get an overview of the content relevant for them in order to be supported in their daily activities.

> *"The IMS contains so much information you don't know where to start looking" (Workshop participant, March 31, 2015)*
> *"If you are not appealed by the front page [of the IMS on the intranet], you do not look at it, and finally you do not search for information there" (Manager, Finance and Control, February 09, 2015)*
> *"Unified notation and frameworks to present information in the system are important so that the employees can navigate." (Workshop participant, March 31, 2015)*
> *"An information-based system would make it easier to use in the organization." (Workshop participant, March 31, 2015)*

The examples and citations above show how the design of the IT artefact, the intranet, in the government agency, shapes the user perceptions about the content of the system and the IMS, and also its use or non-use. The respondents also suggested many alternative ways of designing the structure of e.g. documents in the intranet. In this case the graphical user interface was commented upon as being very important, including the possibilities to search for relevant content (e.g. documents on safety or other rules or regulations), the structure of the document linked to particular work processes, etc. Properties regarding design also included for example a uniform appearance that enables recognition, the ability to personalize views based on roles in the organization or finding and defining the appropriate metadata to facilitate filtering. This would make it easier to find the appropriate documents linked to a certain area within the large and heterogeneous organization or linked to a particular task in order to get support.

As stated above we can also interpret the intranet as a symbolic system [34], thereby focusing the information stored in the system. With this perspective, we can identify important features of the intranet carrying the IMS such as a large amount of information that is difficult for users of the intranet to absorb and thereby to get supported by and managed through – one of the most important roles for an IMS. At a workshop the participants discussed the possibility of moving from the existing document-based intranet to an IMS carried by an information-based system. This was thought of as allowing more flexibility and effective communication with different types of content such as images, video and text. Once again, the design of the intranet is critical for the IMS and how it is interpreted and used within the organization. We interpret that this goes beyond the initial intention and scope of the intranet in this case study. This means that the expectations of the intranet effectivity are not met when it comes to design, use and governance intentions from the top management group.

Gosain [16] studies enterprise information systems from an institutional perspective and proposes: "[…] the view of enterprise information systems as carriers of institutionalization. Enterprise information systems represent a particular template for the conduct of the organization's activities. At the core of enterprise software packages are numerous assumptions about how organizational processes should work; but even with careful configuration, the fit to organizational needs is often lacking [37]. As the organization evolves over time and the assumptions built into a system's configuration stay static, this misfit has the potential to increase." (ibid., p. 168). The point of departure in this analysis of the *intranet as a carrier of institutionalization* also shows that the IMS as the content is a template of conduct. However, not as active as it could

be regarding the legitimation of it by managers not carrying it optimally (above) and the overall design and functions in the intranet not having a good usability and representation – and therefore not optimally carried by the intranet. This means that several important dimensions are not supported well enough by the intranet as the sole carrier the IMS. For example, the regulative elements are not carried enough, supporting actions in line with specifications and rules. Normative dimensions [34] in terms of standards are not communicated well enough using the IMS as a vehicle to support and govern daily work in the government agency. This also goes for the symbolic values that are expected to be a part of the IMS and the intranet in order to achieve the governance expected. As stated by e.g. [16, 19, 37] this line of reasoning can also be associated with achieving fit between the behaviour, processes and structures in the organization and the design of the systems (the IMS and the intranet) as such longitudinally – a snapshot of fit is not enough.

4.3 Findings and Lessons Learned – How the IMS Is Carried

Table 1 below summarizes the intentions, the present situations and the needs identified in the analysis above focusing on the IMS. The findings clearly show that the IMS could be carried more distinct and be made more legitimate by the management group and the intranet respectively in order to be effective in the organization and to govern the structure, processes and culture. The intentions are not fulfilled and several needs are identified in order to achieve a legitimate IMS governing the organization (Table 1).

A lesson learned from the findings in the case study is that the vision of having one integrated IMS can be questioned using an institutional perspective including *path dependency* [41]. One integrated management system seems to be very, or even far too, complicated to be carried by actors and artefacts (e.g. an intranet) in a large and heterogeneous organization, so the type of and level of integration is challenging [18, 43]. An aligned [36] structure and one IMS or several interlinked IMSs seem to be a critical choice regarding the challenges and relative success of management and control of behaviour in terms of relational systems, activities, and artefacts. Mirroring the findings in the present case study with Jørgensen et al.'s [18] three levels model of integrating management systems, the case is intended to be "coordinated and coherent" (model B).

However, in practice the *misaligned* (Grembergen [13]) IMS is not effective and can be even counterproductive. An IMS can be aligned or misaligned in relation to the three dimensions of structure, process and people [19]. Achieving an aligned and legitimated IMS is therefore crucial – an IMS that is *legitimated* through actors and artefacts as combined *carriers* (cf. [33]). In order to have an effective IMS the case study has also shown that it is important to use the potential of the directive power of an intranet carrying the IMS. Using an intranet as an important and aligned carrier of central pillars when governing an organization, has a potential in the studied government agency. However, this demands a well-designed, well-structured, legitimate and embedded artefact, containing usability for different roles and actors within a heterogeneous organization.

Table 1. Findings summarized.

Analytic theme	Intention	Present situation	Needs
The IMS	One integrated IMS to merge and govern two previously different (separate) agencies	Formulated in strategy, but not achieved in practice (e.g. normative [values and expectations] and cultural-cognitive elements not coherent)	To clarify the institutional setting of the IMS and to create one IMS in strategy and practice or to develop several dialects of an IMS in order to align the IMS with the organization and the intranet (below)
The management group carrying the IMS	An actively and legitimately carried IMS	The IMS not legitimated enough by the management group symbolically, relationally and activity wise	To actively legitimate and carry an IMS in order to strengthen norms and values
The intranet carrying the IMS	An actively and legitimately carried IMS	An intranet design not supporting (not using appropriate cognitive and normative elements embedded in the arte-fact) and effectively carrying the IMS when creating a shared understanding	A redesign of the intranet to support the IMS focusing the structuring of information, design, reach and range in order to make use of the directive power in the artefact

The analysis of the management group at the government agency and the intranet as carriers of important elements of institution also illustrates the importance of *combining and aligning carriers* (in terms of e.g. relational systems [management group] and the artefact [the intranet carrying the IMS]) when trying to realize the intentions embedded in an IMS. Non-aligned and non-legitimate (misalignment according to Gosain [16]) norms, values and behaviour, for example, do not lead to a successful use of an IMS. Misalignment is shown in the case in three dimensions: (1) within the IMS as such and (2) by the management group (business) and (3) IT (the intranet) carrying the IMS. Applying an institutional theory perspective and the focus on carriers have been generative when understanding the IMS in the government agency (further elaborated below).

5 Conclusions

The purpose of this paper has been to investigate, show and reflect upon how institutional theory – focusing on institutional carriers with the support of pillars and path dependency – can help us to understand how an IMS is represented through human

actors and technology in a government agency. The conclusion of this research is that the application of institutional carriers [34] as a perspective on IMSs can help us to explore and understand, in our case, a management group and an intranet as an example of an IT artefact, the role and the challenges as well as relative success of such systems. It is evident that the institutional perspective can help us to understand the symbolic and relational systems that come into play in an organization using an IMS and the human actors, routines and artefacts linked to regulative, normative and cultural-cognitive pillars within. The findings and lessons learned from the case study also lead us to the conclusion that the interplay between the general intention behind an IMS and significant actors and artefacts carrying it are highly important when governing a government agency and its structures, processes and culture. This finding is original in the context of IMS, applying institutional theory when studying human actors and an IT artefact as carriers and is one important contribution from this paper.

The conclusions show the strength of institutional theory, in terms of path dependency, when exploring IMSs – linking the present to the past [34, 41]. Addressing path dependency in this case made it possible to understand the importance and the attention that the agency management gave the aspect of having one integrated IMS – instead of e.g. heading for a more manageable and more effective IMS package (cf. [22]). It is also linked to how IT in general is governed, where the intranet in this case is an example [13, 19]. One potential weakness of institutional theory is that the theoretical perspective can be too dominating, using distinct theoretical concepts, seeing only what the theory (as a lens) prescribes when analyzing the empirical material (cf. [42]). In the case study, however, we have tried to avoid that by using a reflexive research approach, with clear inductive parts, interpreting the empirical material more openly in the early stages, as stated in the research approach section.

The implications of this study is that further research and practice should pay more attention to institutional carriers when studying and improving IMSs. Another implication is the potential to further address human actors and artefacts carrying IMS. There is also a need to deal with the limitation that we have so far only studied one government agency in one national setting. However, we believe that the current case study is a generative example to learn from when focusing IMS and carriers, but there is a need to broaden the empirical data to also include other contexts.

Acknowledgements. This study is partially financially supported by anonymized government agency.

References

1. Alvesson, M., Sköldberg, K.: Reflexive Methodology: New Vistas for Qualitative Research, 2nd edn. SAGE, London (2009)
2. Axelsson, K., Melin, U., Lindgren, I.: Public e-services for agency efficiency and citizen benefit - Findings from a stakeholder centered analysis. Gov. Inf. Q. **30**(1), 10–22 (2013)
3. Balocco, R., Ciappini, A., Rangone, A.: ICT governance: a reference framework. Inf. Syst. Manag. **30**(2), 150–167 (2013)

4. Bannister, F., Connolly, R.: ICT, public values and transformative government: a framework and programme for research. Gov. Inf. Q. **31**(1), 119–128 (2014)
5. Baptista, J., Newell, S., Currie, W.: Paradoxical effects of institutionalisation on the strategic awareness of technology in organisations. J. Strateg. Inf. Syst. **19**(3), 171–183 (2010)
6. Barley, S.R.: Technology as an occasion for structuring: evidence from observations of CT scanners and the social order of radiology departments. Adm. Sci. Q. **31**(1), 78–108 (1986)
7. Berger, P.L., Luckmann, T.: The Social Construction of Reality - A Treatise in the Sociology of Knowledge. Anchor Books, Random House Inc., New York (1967)
8. Bernardo, M., Simon, A., Tarí, J.J., Molina-Azorín, J.F.: Benefits of management systems integration: a literature review. J. Clean. Prod. **94**, 260–267 (2015)
9. Boell, S., Cecez-Kecmanovic, D.A.: Hermeneutic approach for conducting literature reviews and literature searches. Commun. AIS **34**(1), 257–286 (2014)
10. Brown, A., Grant, G.: Framing the frameworks: a review of IT governance research. Commun. AIS **15**, 696–712 (2005)
11. Currie, W.: Contextualising the IT artifact: towards a wider research agenda for IS using institutional theory. Inf. Technol. People **22**(1), 63–77 (2009)
12. Currie, W.L., Swanson, E.B.: Special issue on institutional theory in information systems research: contextualizing the IT artefact. J. Inf. Technol. **24**(4), 283–285 (2009)
13. De Haes, S., Van Grembergen, W.: An exploratory study into IT governance implementations and its impact on Business/IT alignment. Inf. Syst. Manag. **26**(2), 123–137 (2009)
14. Fountain, J.E.: Building the Virtual State. Information Technology and Institutional Change. Brookings Institution Press, Washington, D.C. (2001)
15. Gianni, M., Gotzamani, K.: Management systems integration: lessons from an abandonment case. J. Clean. Prod. **86**, 265–276 (2015)
16. Gosain, S.: Enterprise information systems as objects and carriers of institutional forces: the new iron cage? J. AIS **5**(4), 151–182 (2004)
17. ISO (International Organization for Standardization): Management system standards. https://www.iso.org/management-system-standards.html. Accessed 16 Mar 2018
18. Jørgensen, T.H., Remmen, A., Mellado, M.D.: Integrated management systems – three different levels of integration. J. Clean. Prod. **14**(8), 713–722 (2006)
19. Ko, D., Fink, D.: Information technology governance: an evaluation of the theory-practice gap. Corp. Gov. Int. J. Bus. Soc. **10**(5), 662–674 (2010)
20. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology, 2nd edn. SAGE Publications, Thousand Oaks (2004)
21. Luna-Reyes, L.F., Gil-Garcia, R.J.: Using institutional theory and dynamic simulation to understand complex e-Government phenomena. Gov. Inf. Q. **28**(3), 329–345 (2011)
22. Malmi, T., Brown, D.A.: Management control systems as a package – Opportunities, challenges and research directions. Manag. Account. Res. **19**(4), 287–300 (2008)
23. March, J., Simon, H.: Organizations, Graduate School of Industrial Administration, Carnegie Institute of Technology. Wiley, New York (1958)
24. Mintzberg, H.: Structure in Fives: Designing Effective Organizations. Prentice-Hall, Upper Saddle River (1983)
25. Myers, M.D.: Qualitative Research in Business & Management. SAGE Publications, Thousand Oaks (2009)
26. Myers, M.D., Newman, M.: The qualitative interview in IS research: examining the craft. Inf. Organ. **17**(1), 2–26 (2007)
27. Orlikowski, W.J.: The duality of technology: rethinking the concept of technology in organizations. Organ. Sci. **3**(3), 398–427 (1992)

28. Orlikowski, W.J., Barley, S.R.: Technology and institutions: what can research on information technology and research on organizations learn from each other? Manag. Inf. Syst. Q. **25**(2), 145–165 (2001)
29. Powell, W.W., DiMaggio, P.J.: The New Institutionalism in Organizational Analysis. University of Chicago Press, Chicago (1991)
30. Robey, D., Boudreau, M.: Accounting for the contradictory organizational consequences of information technology: theoretical directions and methodological implications. Inf. Syst. Res. **10**(2), 167–185 (1999)
31. Rose, J., Persson, J.S., Heeager, L.T., Irani, Z.: Managing e-Government: value positions and relationships. Inf. Syst. J. **25**(5), 531–571 (2015)
32. Scott, W.R.: Institutions and Organizations, 2nd edn. SAGE, Thousand Oaks (2001)
33. Scott, W.R.: Institutional carriers: reviewing modes of transporting ideas over time and space and considering their consequences. Ind. Corp. Change **12**(4), 879–894 (2003)
34. Scott, W.R.: Institutions and Organizations, 4th edn. SAGE Publications Inc., London (2014)
35. Simon, A., Karapetrovic, S., Casadesus, M.: Evolution of integrated management systems in spanish firms. J. Clean. Prod. **23**, 8–19 (2012)
36. Symons, C.: IT Governance Framework. Forrester Best Practices, March 2005
37. Talbert, N.: Getting the most from an enterprise system. MIT Sloan Manag. Rev. **44**(1), 11 (2002)
38. Tallon, P.P., Pinsonneault, A.: Competing perspectives on the link between strategic information technology alignment and organizational agility: insights from a mediation model. MIS Q. **35**(2), 463–486 (2011)
39. Van de Ven, A.H.: Engaged Scholarship – A Guide for Organizational and Social Research. Oxford University Press, Oxford (2007)
40. Venkatraman, N., Henderson, J.C., Oldach, S.: Continuous strategic alignment: exploiting information technology capabilities for competitive success. Eur. Manag. J. **11**(2), 139–149 (1993)
41. Vergne, J.-P., Durand, R.: The missing link between the theory and empirics of path dependence: conceptual clarification, testability issue, and methodological implications. J. Manag. Stud. **47**(4), 736–759 (2010)
42. Walsham, G.: Interpretive case studies in IS research: nature and method. Eur. J. Inf. Syst. **4** (2), 74–81 (1995)
43. Wilkinson, G., Dale, B.G.: An examination of the ISO 9001:2000 standard and its influence on the integration of management systems. Prod. Plan. Control **13**(3), 284–297 (2002)

Live Enrolment for Identity Documents in Europe

Tarmo Kalvet[1(✉)], Henrik Karlzén[2], Amund Hunstad[2], and Marek Tiits[3]

[1] Tallinn University of Technology, Akadeemia tee 3, 12618 Tallinn, Estonia
tarmo.kalvet@ttu.ee
[2] Swedish Defence Research Agency, 16490 Stockholm, Sweden
[3] Institute of Baltic Studies, Lai 30, 51005 Tartu, Estonia

Abstract. Digital image alterations (morphing) of identity document photos is a major concern and may potentially allow citizens with malicious intent to enrol for identity document(s) later to be used also by another individual. Taking the photo in the application office – live enrolment – can address this issue. However, this is a break with tradition and entails a sizeable overhaul in the public sector, which can be reluctant to change and often lacks the necessary formal methods that ensure a smooth transition. The objective of this paper is to map the main barriers and drivers related to live enrolment based on theoretical research and interviews conducted with high-ranking officers at passport authorities in Estonia, Kosovo, Norway and Sweden. These countries have successfully switched to live enrolment. The main driver for live enrolment has been increased security; for Estonia, user convenience was important and was behind the decision of keeping alternative application processes for the citizens around. The absence of legacy systems makes it easier to implement public sector innovations, such as live enrolment. Behind the successful implementation is proper risk management, covering technological, political and organisational risks. Finally, the research results indicate varying experiences, obstacles, cultural differences and trade-offs, emphasizing the need to understand barriers and drivers in a contextualised way.

Keywords: Morphing · Identity document · Passport · Live enrolment
Social acceptance · Public sector innovation · Drivers and barriers
Risk management

1 Introduction

Existing facial recognition algorithms are unable to fully deal with digital image alterations. This means that malicious citizens are able to apply for identity documents that can be used also by another individual. When a photo is brought in the photographer, the applicant, and others involved, all must trust that the photo has not been tampered with. In theory, biometric facial recognition algorithms should be able to determine altered photos from genuine ones, but, in practice, the algorithms are not perfect.

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 29–39, 2018.
https://doi.org/10.1007/978-3-319-98690-6_3

While the live enrolment of fingerprints is a common standard for identity documents, such as European passports, this is not the case for passport photos. The non-live photo enrolment procedure is vulnerable to 'morphing attacks' [1], in which a digitally altered photograph is enrolled in order to allow two or more persons of similar appearance to use the same passport to pass visual checks of their face (both automated and manual) at border crossing points. This would mean that a known criminal could travel with the passport of somebody else. More generally, if one does not make sure that the photo is an authentic representation of the person applying for the passport (and nobody else!), the photo will not provide the intended security. For instance, someone may be forced to apply for a passport with somebody else's photo, resulting in identity-theft where that somebody could travel in the name of the coerced. This could be useful for a criminal involved in e.g., illegal immigration or human trafficking. The adoption of a common standard for live enrolment of passport photos may eliminate, or at least greatly reduce, these threats.

As of 2016, only some European countries have switched to live enrolment (Fig. 1).

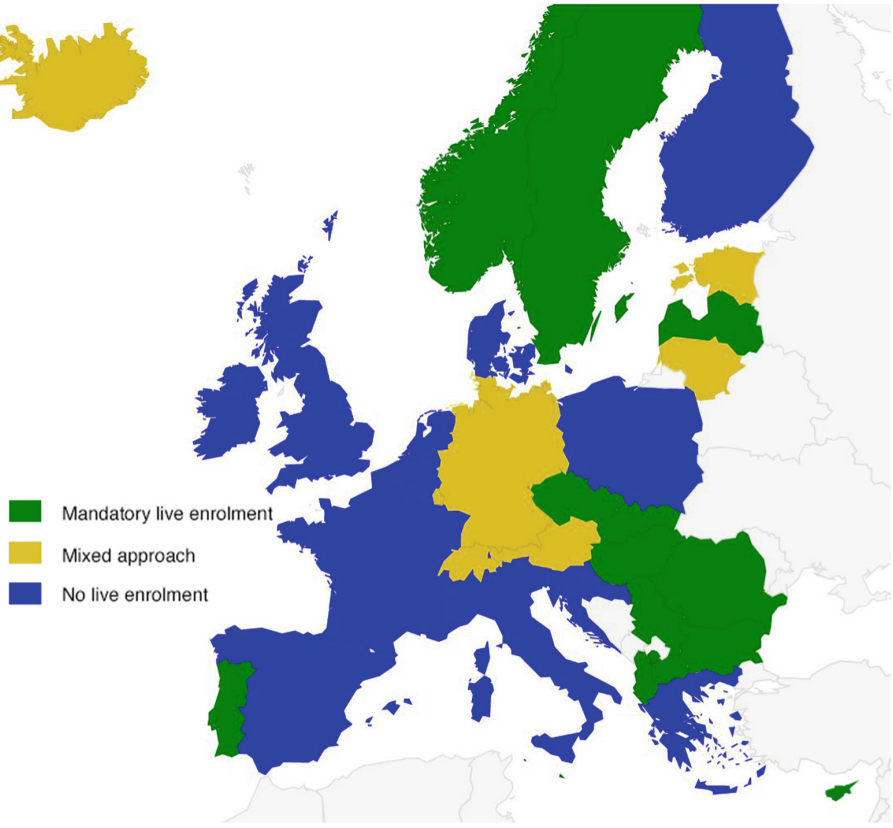

Fig. 1. State of the live photo enrolment in the Europe Source: Authors on the basis of national experts participating in the "Sixty-sixth meeting of the Committee on a uniform format for visa" (16 March 2016, Brussels) and on national passport application websites.

Some countries are applying a mixed approach where live enrolment is offered as an option, while most of the European countries are still applying traditional approach where applicants bring their photos with them. Primarily Northern and Southeastern European countries have made live photo enrolment mandatory. Meanwhile, a number of larger member states, such as France, Italy, Spain and the UK, have not yet launched live enrolment. In some cases, the existence of live enrolment varies at the intra-country level; for instance, it is available only in a third of Swiss cantons and various German municipalities.

The objective of the paper is to explain the main drivers and barriers behind switching to live enrolment. These are vital to understanding the problem at hand, its possible solutions, and to propose policy recommendations to other countries. It is important to note that there might not be one best model suitable for all countries as cultural and economic differences must be considered.

The research is carried out on the basis of four countries – Estonia, Kosovo (Kosovo is recognised as a country by most, but not all, EU member states), Norway and Sweden – in the issuance of passports. These countries represent a variety of practices. In Kosovo, Norway and Sweden live enrolment is mandatory. Estonia, on the other hand, represents a country where a mixed approach is in place. These countries represent an economically, socially, and culturally diverse group of countries, therefore they were selected as case study countries for this research.

As the first step, a review of existing academic and policy literature was conducted to search for influential factors that may affect the implementation of live enrolment. Since the specific topic is scarcely discussed in literature, the literature search was expanded to involve academic papers and policy reports on public sector innovation and technology acceptance. As a result, an inventory of potential barriers and drivers was compiled, covering technological issues, organizational, user aspects and other relevant factors.

For the second step, interviews were carried out with high-ranking persons at passport authorities of each country. Interviews were either conducted in person or by phone. The respondents – all government employees with a top-level passport process responsibility (currently or previously) – were selected to be knowledgeable and have good insight into the matter at hand.

The paper is structured as follows. In section two, a general overview of the live enrolment processes in Sweden, Norway, Kosovo and Estonia is provided. This is followed by literature review. In Sect. 4 the formulation of several hypotheses about challenges and experiences of full live enrolment based on the theoretical studies and previous empirical work takes place. Section 5 discusses the findings, followed by the conclusions (Sect. 6).

2 Live Enrolment Processes in Sweden, Norway, Kosovo and Estonia

In Sweden there is only one way of applying for a passport, and it is in person at a passport application office where a facial image is taken. This includes taking a digital photo and it is not possible for the applicant to receive a copy of the photo. The system was introduced in 2005.

Norway's process is highly similar to the Swedish one and was also introduced in 2005.

In Kosovo live enrolment is also in place, a digital photo has been taken close to the table of the clerk since 2008.

Estonia, an example of a typical mixed enrolment process, accepts three different modes of the submission of passport photos since 2007:

- Live enrolment: Passport application with a photo taken in special photo booth, which links the digital photo with the personal identity code of the person. An important difference to e.g. the Swedish case is that the photo booths are self-service rather than operated by application office staff. The differences may have implications for both quality and security.
- Traditional application for passport on a paper along with printed photo, submitted in person in Estonian Police and Border Guard Board or sent in by postal mail;
- Electronic passport application (filled in PDF form) together with photo in JPG format in the same electronically signed document (BDOC).

3 Literature Review

Studies on ICT-driven innovation in the public sector frequently emphasize the organizational, administrative and political context as a source of innovation drivers and barriers. Since the application of the live enrolment requires changes in organizational workflows and coordination between different organizations, it faces a number of organizational barriers. Such barriers involve, for example, existing governmental silos and lack of communication, the complexity of organizational change, and concerns about high implementation costs.

Studies have found that the innovation capacity of public sector organizations depends on a number of organizational factors, such as organizational structures, intra- and inter-organizational collaboration and coordination, organizational culture, leadership styles [2, 3], resources [4], qualified personnel [5], etc. Public sector organizations are also strongly affected by the legal culture and administrative traditions of the state [4]. Moreover, the inherent complexity of public sector organizations and their accountability to a multitude of stakeholders make the implementation of organizational changes much more difficult in the public sector than it is in the private sector [6].

Although the public sector context generates more barriers than drivers to innovation, some drivers still exist. Firstly, drivers may be generated by external triggers, such as competitive pressure by other organizations, countries or international bodies [4], legal obligation, political priority and public demand [2]. At the organizational level, innovation can be driven by participation in cross-organizational and cross-border knowledge transfer networks [7], and strong leadership by administrative and political managers [4].

The regulatory and legal context is perceived to be important in several respects. On the one hand, existing regulations often stifle innovation; on the other, regulations

can also promote innovation, for example by imposing a legal obligation to implement certain solutions [2].

Another potentially important factor is demand by citizens and businesses and can act as an influential factor for live enrolment. In order to understand what factors affect the demand and acceptance of live enrolment, research on the acceptance and use of technology can provide valuable insights. The general point of departure of such literature is the understanding that there are a number of factors that influence the user as to whether or not to adopt a novel technology. One popular approach for mapping those factors is the technology acceptance model (TAM), which argues that acceptance is determined by the perceived usefulness and ease of use of a technology [8]. TAM's derivative, the Unified Theory of Acceptance and Use of Technology (UTAUT) is more elaborate and incorporates additional factors, explaining how a decision is formed about the use of an information system. The theory builds on four key constructs: (1) performance expectancy, (2) effort expectancy, (3) social influence, and (4) facilitating conditions [9, 10].

4 Hypotheses

On the basis of the conducted theoretical research and previous empirical work, we have formulated several specific hypotheses about barriers and drivers of full live enrolment. This includes aspects pertaining to both the users – individuals expected to use the live enrolment service – as well as the suppliers of said public service.

Our key hypotheses are related to the expectations of the government when introducing live enrolment. We expect security related benefits, such as making sure the photo comes from the person who is applying for a passport and is not altered in any way as well as other kinds of protection from document forgery. In addition, central arguments around introducing live enrolment relate to the accuracy and reliability of the identification of persons and protection from identity theft. We also expect to see more general public security policy objectives, such as the fight against illegal immigration and human trafficking, as well as the fight against terrorism and serious crime [11, 12].

We expect the general key barriers and drivers of public sector innovation to be the same for the specific case of live enrolment, with the importance of continuous political support to the process, financial resources, commitment and leadership of administrative and technical managers ("championing of the project"), ICT awareness and capabilities of the stakeholders (systems, skills, tools and methods) [2, 3, 13].

Existence of a suitable live enrolment technology on the local market, or suppliers who can provide it, might also be important as well as related business models. The cost of implementing live enrolment may also be mitigated by similar existing systems and solutions. If live enrolment for ID cards is already in place, the step towards live enrolment for passports should be considerably easier. Extensive population censuses (registries) can further help as there will be more information in the system already. Conversely, going for live enrolment could also be made easier if other changes were anyway needed to the passport application process, such as introducing fingerprint scanners.

Any security measure must be cost-efficient and in some cases it may be that population density (e.g. in relation to the number of application offices) is too low to offset the cost of live enrolment. On the other hand, live enrolment could be a more cost-effective security measure than e.g. ones based on complex certificate distribution, or instead less cost-effective than putting a limit to the amount of new passports a citizen is allowed in a set period of time.

As public sector innovation is risky, we expect to see evidence of resistance from some stakeholders (e.g., agencies reluctant to take on additional tasks). This may be particularly likely in case procedures were recently changed. Related to this is the use of formalized methodological approaches to implementing the change (a clear definition of work rules and methods, training, etc.) in order to mitigate the risks. Since general experience in public sector innovations is that only limited attention is paid to precise systematic gathering and use of measurement and data, we do not expect to see very clear measurable target indicators set in relation to live enrolment.

There are also certain probable expectations from the public which can influence live enrolment uptake [14]. Convenience of the collection of photos (e.g. easier to use photo booth than make the extra effort to a professional photographer) and image satisfaction are likely important factors. Furthermore, the spread of digital cameras could have a two-sided effect: on the one hand, encouraging uptake of related technologies such as live enrolment, while on the other hand, resistance from professional photographers due to an increasingly smaller market may appear. We also expect to see some problems due to the innovative nature of technology. Societal groups such as those less familiar with technology and digital photography may be more risk-averse and resistant to live enrolment, preferring the traditional method; age and profession distribution of a region or country may be underlying factors here. Pre-existing widespread use of privacy-sensitive technology, like biometrics and databases, may make live enrolment easier to accept. For instance, fingerprint usage has previously had bad connotations due to the connection with criminal records, but this has been alleviated by the rising use of fingerprint logins on smartphones.

Finally, since the main purpose of passports is to travel to other jurisdictions, a country must consider not only its own perceptions but also external pressure. For instance, countries must take into account demands from the EU and International Civil Aviation Organization (ICAO) as well as particular countries like the US requiring certain procedures to ensure visa-free entry.

These specific hypotheses were the main topics focused of the interviews and other data collection, and in the following section key findings are discussed.

5 Discussion

If current facial recognition algorithms are unable to fully deal with digital image alterations, and live enrolment– where photos are taken in a controlled environment in the application office – provides a way of alleviating this risk, then the question becomes why are all countries not already using it. Our research shows that some countries have switched to live enrolment successfully, while many still use the

traditional method of having citizens bring photos when applying for a passport. Also, many countries allow both methods.

Building on the theories of public sector innovation and acceptance and use of technology, the authors have identified a number of factors influencing live enrolment decisions and implementations. These factors may be split in those behind the decision to switch to live enrolment and the factors impacting the ease of implementation.

5.1 Making the Decision to Go Live

A main driver for live enrolment is increased security, hindering false images in passports. Furthermore, live enrolment is convenient for citizens as they do not need to first acquire photos before visiting the application office. Also, formatting issues are unlikely to occur if all photos are taken by the same organisation. However, all security measures must be weighed against the different associated potential costs. For instance, the passport authority will need to administer photos in a different way, while portrait photographers lose a fairly substantial part of their income. Photos not taken by experienced photographers may also lead to less flattering portraits, which may be associated with a decrease in satisfaction among citizens. Some photographers have indicated that they believe they could provide as secure an image process as the government officials and this may ultimately be a question of trust.

There may also be old connotations of feeling like a criminal attached to biometrics being captured by the authorities and care must be taken to give applicants sufficient privacy, especially for those with certain religious beliefs or disabilities. Indeed, the passport authorities in Sweden and Norway are both trying to be service minded and create good will.

Another factor affecting the decision of switching to live enrolment is the external global pressure from organisations like the EU and ICAO, as well as influential countries like the USA who require certain procedures for visa free entry. In general, however, EU member states seem to be doing very well with passport security in an international context. There may be pressure coming from other countries having implemented live enrolment or incentives coming from other parts of one's own government who use similar technology. On the other hand, traditions may be strong and if a country recently changed its application procedures it will likely be more reluctant to do so again.

It is also interesting to note that the motivation for going live has been rather different in our case study countries. Estonian prioritised the quality of photos, as the photos were previously not always of sufficient quality, while security considerations, such as avoiding photo morphing were not of concern. Contrastingly, for Kosovo, Sweden and Norway security considerations were the highest. For Estonia, user convenience has been very important and is behind the decision of keeping alternative application processes open for the citizens.

Another finding is that in the absence of legacy systems (like in Kosovo), more profound decisions are generally reached (and implemented) in an easier fashion. Also, it must be reiterated that decisions are easier to reach if there are other organisations that have introduced live enrolment for related services – it is possible to learn from

their experience and somehow pressure to innovate is created from the public side if they have positive experience with live enrolment in other context.

In general, still, decisions to employ live enrolment seems to be influenced equally both by the context external to public sector organizations as well as by organisational level factors. External pressures alone (such as security related considerations by politicians) cannot explain the decisions to switch, as more countries would have switched to live enrolment by now. Suitable organizational level context (including supportive organisational culture, allocation of resources), coupled with individual level drivers of key persons (such as job-related knowledge and skills, willingness to exploit risky avenues) behind the live enrolment introduction, were also needed.

5.2 Implementing Live Enrolment

Even if the decision is made to introduce live enrolment, proper risk management is needed to account for any and all setbacks encountered during implementation. One possible risk is that appropriate technology is missing or that vendors do not have suitable offerings. There are not many vendors available and, as such, there is low competition that may have led to higher prices and doubtful quality. There are ideas of certifying vendors to address this. One of the typical problems related to public sector innovation is related to the lack of suitable technology in market. Thus, there is need to develop related technologies further until they can be applied. This also calls for the implementation of rather unpopular public procurement mechanism, so-called public procurement of innovation. In the case of live enrolment suitable solutions did exist in the global market. However, it should be noted that any vendor winning a particular procurement process will effectively have a monopoly for live enrolment in the applicable region and time of that procurement. As such, there is considerably more at stake than in the case of non-live enrolment where there can be many simultaneous vendors sharing the market. Thus, care must be taken in order not to let corruption distort the procurement process.

Furthermore, staff must be trained. In an unstable political situation, a decision may soon be reversed and idea championing and motivation could be limited. Overall, a culture rife with innovation will be more accustomed and prepared for changes and therefore more likely to succeed with a change in enrolment processes. Of special concern is when passport applications are done for extraordinary circumstances, such as when conducted outside of one's country, when applying for temporary/emergency passports, or when applying for visas. Indeed, citizens seem to highly value the speed of the application process, with examples of citizens making sure they get passports with the old ten year validity before a switch to a mere five year validity. In general, citizens seem more concerned about the use of fingerprints and biometric data in databases than about live enrolment. Passport officials have suggested changing the discourse and focusing on the improved security measures to stop identity theft.

Our research confirms that one of the typical weaknesses associated with insufficient attention to the use of impact assessments and evaluations is also true regarding live enrolment. None of the countries set target indicators nor was their achievement monitored. Neither have we encountered (public) impact assessments of the benefits of

live enrolment (that in turn might slow down other countries to switch as the benefits are debatable).

Our research of technology acceptance aspects did not reveal problems. The overall societal context was supportive as people had already experience with digital photography and automated service machines generally.

5.3 A Multi-faceted Situation

The interviews revealed varying experiences, obstacles, cultural differences and trade-offs. Only in Sweden was the voice of the photographers an issue. Indeed, Estonia may have found a move towards live enrolment easy since they also kept the traditional application method. However, Norway also went full live and in fact had a lot of political support and funding, although the support and funding varied over time. Kosovo, being a rather young country, seems to have succeeded with implementing live enrolment thanks to no pre-existing alternatives. While there were some discussions among photographers, their voices were not very strong. Furthermore, the issue was not particularly political, in part thanks to a small government. Also, since Kosovo's efforts started already in 2000, there was no issue with the advent of digital cameras further affecting photographers.

The other parts of the identification chain were also important influencers. The existence and levels of birth registries vary widely in the EU, as do the use of automatic border gates, mobile application kiosks and restrictions on citizens' right to a passport. The use and sophistication of other forms of ID may also play a role. For instance, the Swedish transport authority previously tried but failed to implement live enrolment for driver's licenses. Now they plan to revisit this with the experiences of the passport authority.

6 Conclusions

Using live enrolment for passports, where photos are taken in the application office, is a topical issue since it can limit the possibility of malicious digital image alterations [15]. However, live enrolment has both its advantages and disadvantages and a trade-off is expected. Live enrolment may make passport applications a one-stop process, ensure consistent formatting and increase security. There may also be spill-over effects. For example, Estonia first introduced live enrolment for driver's licences, and this experience simplified introduction of the live enrolment for passports. Sweden is following a reversed path from live passport enrolment making the similar move for licences simpler. External pressure can also be international e.g. from the ICAO or EU.

However, live enrolment is a break with tradition and depends on overhauls in the public sector where measurable targets and risk management is typically wanting and the passport authority may be reluctant to take on new tasks. The public may hesitate to trust the government further with their private data and find that their portraits are now less flattering. Furthermore, live enrolment can make it more difficult for expats, who need to apply for a passport from abroad. Furthermore, a lack of competition among live enrolment equipment vendors may lead to high prices for low quality, which both

Sweden and Norway have found. Portrait photographers meanwhile suffer loss of income.

To minimise the trade-off, authorities have tried becoming more service-minded, creating good will, not least in relation to the speed of the application process, which is highly important to applicants. Portrait photographers, on the other hand, have suggested that they be certified, in a similar vein to what is done in other areas where the private sector provide a service for the public good, thereby ensuring adequate levels of security. However, at least so far their suggestions have not become a reality. In the case study of four countries, only in Sweden was the voice of the photographers at all an issue and even there the authorities chose to prioritise security and live enrolment.

The decision to introduce live-enrolment and the state of live enrolment in Europe are both multi-faceted. Experiences, drivers and obstacles vary between countries and sometimes there are even more regional structures where one part of a country differs from another. Furthermore, it is not a binary decision: there are different versions of live enrolment implementations and also mix enrolment where live enrolment is offered but not mandatory, such as in Estonia. What is more, when planning the introduction of live enrolment, Estonia prioritised quality of photos and user-friendliness while Sweden focused on security. Implementation wise, Kosovo benefitted from a small government and no pre-existing alternatives, while also pre-empting the digital camera era which later put more pressure on photographers. Norway had substantial political support and funding, although the backing varied over time, while Estonia utilised Schengen funds.

As such, the decision to introduce live enrolment and successfully implementing it, is dependent on a vast number of cultural and political factors. Straightforward sociocultural models are unable to fully explain the current situation and the overall state of a country's passport maturity with its automatic border gates, mobile application kiosks, restrictions on passport renewals as well as the proliferation of birth registries and the security of breeder documents.

Several limitations remain, however. First, the attitudes of photographers (as important stakeholder group) could be studied further. They might be hard to reach, though, as those who suffered economically due to live enrolment could no longer be in business. Also, empirical data could be collected from the countries that do not practice live enrolment now to deepen current results.

Acknowledgments. This work was supported by the European Commission through the project "Fast and trustworthy Identity Delivery and check with ePassports leveraging Traveler privacy" (FIDELITY, grant 284862) funded from FP7 and a grant "Public sector innovation: The case of modern identity management technologies" (PUT773) from Estonian Science Foundation.

References

1. Ferrara, M., Franco, A., Maltoni, D.: The magic passport. In: IEEE International Joint Conference on Biometrics (IJCB), pp. 1–7 (2014)
2. De Vries, H., Bekkers, V., Tummers, L.: Innovation in the public sector: a systematic review and future research agenda. Public Adm. **94**(1), 146–166 (2016)

3. European Commission: Powering European Public Sector Innovation: Towards a New Architecture. Report of the Expert Group on Public Sector Innovation (2013)
4. Bekkers, V., Tummers, L., Voorberg, W.: From public innovation to social innovation in the public sector: a literature review of relevant drivers and barriers. Erasmus University Rotterdam, Rotterdam (2013)
5. Alshehri, M., Drew, S.J.: E-government principles: implementation, advantages and challenges. Int. J. Electron. Bus. **9**, 255–270 (2011)
6. Rashman, L., Withers, E., Rashman, J.H.: Organizational learning and knowledge in public service organizations: a systematic review of the literature. Int. J. Manag. Rev. **11**(4), 463–494 (2009)
7. Albury, D.: Fostering innovation in public services. Public Money Manag. **25**, 51–56 (2005)
8. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
9. Venkatesh, V.: Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf. Syst. Res. **11**(4), 342–365 (2000)
10. 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)
11. Hoepman, J.-H., Hubbers, E., Jacobs, B., Oostdijk, M., Schreur, R.W.: Crossing borders: security and privacy issues of the European e-passport. In: 1st International Workshop on Security, Kyoto, Japan, 23–24 October (2006). http://arxiv.org/abs/0801.3930
12. Tiits, M., Kalvet, T., Laas-Mikko, K.: Social acceptance of epassports. Lect. Notes Inform. **P-230**, 15–26 (2014)
13. Kalvet, T.: Innovation: a factor explaining e–government success in Estonia. Electron. Gov. Int. J. **9**, 142–157 (2012)
14. Kalvet, T., Tiits, M., Laas-Mikko, K.: Public acceptance of advanced identity documents. In: Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2018), Galway, April 2018
15. Behrensen, M., Kalvet, T., Kindt, E., Tiits, M., Palm, E.: Harmonization of epassport and border checks in the European union: ethical, social, and legal aspects. A summary of the practical implications of the legal, ethical, and social analysis of the FIDELITY project (2016)

Understanding Public Healthcare Service Quality from Social Media

Hong Joo Lee[1], Minsik Lee[2], and Habin Lee[3(✉)]

[1] Catholic University of Korea, Bucheon, Gyeonggi 14662, Republic of Korea
hongjoo@catholic.ac.kr
[2] Yonsei University, Seoul 03722, Republic of Korea
salvia04l3@yonsei.ac.kr
[3] Brunel University London, Uxbridge UB8 3PH, UK
habin.lee@brunel.ac.uk

Abstract. Despite the opportunities and demands to use social media to support public policy-making processes, a systematic approach to reflect social media sentiments in policy making processes is yet to be proposed in the literature. This paper provides a method to assign tweets into one of SERVQUAL dimensions to identify sentiments and to track perceived service quality for policy makers in national health services (NHS). In this study, we devise a methodology to (1) identify more reliable topic sets through repeated LDA and clustering and (2) classify tweets with the topics based on a theory in service quality. To demonstrate the applicability, we selected healthcare as our target area and picked the NHS of U.K. for sensing the service quality of public policy. We collected tweets about NHS for about 4 years and created dictionaries related to the domain of healthcare with user reviews on hospitals and general practitioners in U.K. We applied the suggested methodology to track social perceptions and compared the applicability among different methods.

Keywords: Social perceptions · SERVQUAL · Healthcare · NHS
Sentiment analysis · Topic modeling

1 Introduction

Extensive amount of online user-generated content or word of mouth have been produced and the surge of its volume is getting accelerated due to social media. Business companies are trying to understand and monitor social perceptions on their brands and products [1]. Firms are doing this by collecting and analyzing user reviews and similar digital traces, including social media, to understand how they are perceived by their communities [2–5]. In addition, social sentiments on certain events can be collected through social media and used to predict outcomes of collective behavior.

A citizen-inclusive approach is increasingly favored by policy makers and ever more robust underpinned by significant amounts of data sets often harvested and available through the internet and social media. Patient experiences shared through social media or online communities include real people talking about what they have experienced and how they feel, in their own words [4]. By listening to online voices, public service

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 40–47, 2018.
https://doi.org/10.1007/978-3-319-98690-6_4

design can be significantly improved [5]. However, due to the large volume of online voices available, it is a challenge to measure social perceptions manually. Nonetheless, there is a big benefit to unravelling the value contained in big data to improve existing public services. Tracking the service quality of National Health Service (NHS) with social media can help us identify dimensions to be investigated for further improvements reducing the number of survey that requires more costs and time. [6] provides an evidence that patient web-based ratings on service experience are associated with hospital ratings derived from a national paper-based patient survey. The analysis of patient stories can be integrated with more quantitative surveys or other technical approaches to provide a comprehensive picture [4]. Despite the opportunities to use social media to support public policy-making processes, existing empirical studies on compiling social media sentiments into service quality measurements for public policy have some limitations in analyzing data and unraveling meanings.

The aim of this study is to devise a method to measure social perceptions on service quality using social media data. For demonstrating the suggested method, we choose healthcare as our application area and select NHS of U.K. to measure service quality of public policy. Since social media data have lots of noisy data, we applied Doc2Vec and machine learning algorithms to identify relevant data on service experience. With the relevant data, we use Latent Dirichlet Allocation (LDA) for getting the results of topic modeling and get more robust topic sets by reiterating LDA and clustering topic sets for lessening subjective bias. We classify tweets with the topics based on an existing framework in service quality – SERVQUAL [7, 8]. SERVQUAL has been widely used for assessing the quality of health services in the literature as its five service dimensions provide policy makers with specific implications for intervention [9–12]. Terms belonging to each topic are matched with words from the pre-classified data and survey questionnaire for each construct of SERVQUAL. In doing so, we can measure similarity values between a tweet and the topic sets. These similarity values are input data for machine learning algorithms to classify a tweet into one of SERVQUAL dimensions and other. Then, a dictionary for the healthcare is built to compute the sentiments. Dictionaries are made from different methods and their accuracies are compared to find the most appropriate one. We collect tweets about NHS for about 4 years, patient reviews on hospitals, and reviews on general practitioners (GPs) to make a dictionary. In addition, we collect survey questionnaires from existing researches on healthcare service quality to match the constructs of SERVQUAL to the topics. With these, we measure social perceptions relating to each dimension of service quality of NHS systematically.

2 Related Work

The starting point of text mining is to extract words from the data and build a term-document matrix. The elements of the term-document matrix are usually term frequency of specific words in a document. Tf-idf is used for the element of the matrix and is the product of term frequency (Tf) and inverse document frequency (Idf) to assess the importance of a term for distinguishing documents [13]. With the term-document matrix, researchers identify the aspects or the sentiments of documents.

For the aspects, basic, stylistic, and semantic characteristics are usually considered [14]. Basic characteristics include information on the document itself, such as posting date, and sentiments of documents. Many studies on measuring sentiments with social media data and online reviews are based on lexicon-based and machine learning approaches [15]. Dictionary-based term matching is one major technique of Lexicon-based approaches and it simply measures sentiments by matching with dictionaries that have predefined sentiment scores for words [16]. By summating sentiment values of a document, we classify whether it is positive/negative or denotes a specific mood. On the other hand, the core of the machine learning approach is training classifiers such as decision tree and SVM with the sentiment-labeled data and apply the trained model to classify unlabeled documents [13].

Stylistic characteristics are related to writing styles that cannot be easily derived by simply browsing—such as the average number of words in a sentence [14], readability, and complexity [17] of documents.

Finally, semantic characteristics are related to the substance of the documents. Some studies such as [18] defined keyword sets to corresponding categories and calculated how many keywords are in a document to assign it to a relevant category. Other studies applied statistical techniques of topic modeling to extract meaning of documents. [14] applied Latent Semantic Analysis (LSA) to identify meanings of user reviews and ordinal logistic regression to classify helpful reviews. LSA applies singular value decomposition to the term-document matrix and extracts the low rank approximation of the matrix [19]. With the reduced matrix, we can understand the meaning of documents within their dimensions. LDA is widely adopted in recent studies to automatically identify latent topics from a collection of documents [20]. LDA is based on the intuition that documents exhibit multiple topics and a topic is "a distribution over a fixed vocabulary" [21]. [22] applied LDA to identify 30 themes within patient feedback and to measure sentiments of all the themes. They provide a better understanding of patient opinion by associating themes and sentiments. [23] applied weakly supervised LDA with the seed words and identified topics according to the SERVQUAL constructs. They selected seed words using only nouns associated with the essence of SERVQUAL dimension and selected these terms directly from the vocabulary of their corpus. They measured sentiments of the constructs and studied its effects to the overall satisfaction rating in online commerce.

Recently, the use of Word2Vec, [24] which represents semantic space of words from very large data set, in studies on text mining and natural language process is increasing. Doc2Vec [25] suggests an unsupervised algorithm that outperform the traditional "bag-of-words" approach in text classification and sentiment analysis with the semantic word representation of Word2Vec. Since Word2Vec and Doc2Vec are not for identifying latent topics, [26] propose a Topic2Vec approach which embed topics in the semantic vector space represented by Word2Vec and compared their result with LDA.

LDA has been adopted in many social science studies for identifying latent topics [27] and is shown better performance than traditional topic modeling methods such as LSA. Other recent approaches including Topic2Vec have not been used in text mining studies. Thus, we used LDA as topic modeling method and suggested our method to track social perception on service quality.

3 Data

Our target public service is the NHS of the U.K. since we can collect sufficient tweets and patient reviews from its website, NHS Choices (http://www.nhs.uk/). We collect 50,716 tweets that contain NHS in their posts from January 1, 2013 to October 31, 2016. We use tweets posted in the U.K. and written in English. We pre-process the tweets by removing URLs, numbers, punctuation marks, stop words, and other languages. Then we extract all words from the tweets and stem the words since one word can have different forms (e.g., pay and paid). We build a term-document matrix with the stemmed words and remove terms with a sparsity greater than 0.9999 to reduce complexity. In addition, the term NHS is removed from the matrix since every tweet contains it therefore meaningless.

Table 1. Top 30 frequent words

Words	Words	Words
Twitter	mp	Tories
Pic	Day	Service
UK	Time	Free
Health	Pay	Private
Care	Labour	England
People	BBC	Money
Staff	Instagram	Privatization
News	Save	Support
Hospital	Patient	Doctor
Trust	Bit	Public

The cell values of the term-document matrix are term frequencies. Table 1 shows top 30 frequent words in the tweets. Though we removed specific URLs for attaching web pages or photos to a tweet, we still have some words including pic, bit and instagram.

4 Methodology

Step 1: Excluding Non-relevant Tweets
Most of the tweets are not relevant to service quality for patients of hospitals or GPs but arguing about healthcare reform, political discussions, NHS budget and so on. Thus, we need to identify tweets about service quality for the further analysis. We apply a machine learning approach to identifying non-relevant tweets. Since we need a training dataset, two graduate school students who are aware of the concepts of SERVQUAL are recruited to classify randomly selected 600 tweets. The purpose of this study and the dimensions of SERVQUAL are introduced to the recruited raters. Two raters first

individually classify the tweets into one of SERVQUAL dimensions and then discuss together to agree on their classification results. The agreed classifications are used as training and test data in the following steps. In the data set, there are more tweets related to Reliability and Tangibles dimensions than other SERVQUAL dimensions though the largest number of the tweets are classified as other.

We also use the survey items of SEVRQUAL studies for the training and test datasets to expand related word lists in corresponding dimensions. We collect survey items of SERVQUAL in healthcare and pre-process them as we did for the tweets in our study.

By doing this, noises from non-relevant tweets can be reduced for performing topic modeling.

Step 2: Repeated Topic Modeling and Clustering

The assumption of LDA is that "documents are represented as a random mixture over latent topics - where each topic is characterized by a distribution over words" [28] and that LDA extracts latent topics among documents. LDA is based on Gibbs sampling which attempts to collect samples from the posterior to approximate it with an empirical distribution. Due to the random selection procedures in the approximation above, the results of LDA vary in different implementations. Researchers choose one set of topics which can explain their data well after repeated trials. [23] use seed words for the five constructs of SERVQUAL to identify corresponding topics through LDA – called weakly supervised LDA [29]. Though they use the seed words to guide their topic selection, it is still grounded on sampling-based algorithm and the selection of the words is done manually.

Unlike the weakly supervised LDA approach, this study runs LDA many times and applies hierarchical clustering to the results of the LDAs for extracting more robust results and for reducing human interventions. This study uses the tweets predicted as relevant in step 1 and runs LDA to have thirty topics with thirty words per topic at one run and reiterate it 1,000 times with varying delta values from 0.1 to 10 [30]. Number of words per topic is usually selected from 20 to 30 and we chose large enough number topics for applying clustering.

As a result of running LDA once, we get 30 topics with thirty words belonging to a topic and their probabilities. With 30,000 topics from 1,000 repetitions, we apply hierarchical clustering algorithm to have similar subsets of topics by calculating the distances of topics with the probabilities of words in a topic.

Step 3: Dimension Classification

We, then, assign each tweet to one of SERVQUAL constructs or to other dimension. Similarity values of each tweet presented in the term-document matrix to the 30 topic clusters described in Step 2 are measured with the Jaccard index. The Jaccard coefficient calculates similarity between finite sets and is defined as the size of the intersection divided by the size of the union of the comparing sets. These 30 similarity values of a tweet to the 30 topic clusters are input values for machine learning algorithms to classify each tweet into one of SERVQUAL or other dimensions. We apply diverse machine learning algorithms and conduct 5-fold cross validation with the labeled 600 tweets as explained in Step 1.

Step 4: Dictionary Building
To use the dictionary-based matching approach for measuring sentiments of a statement, we need a dictionary which has sentiment values of words. AFINN [31] assigns words with negative scores for negative sentiment and positive scores for positive sentiment. Bing [32] and NRC [33] categorize words in a binary fashion into positive or negative category. Since the widely used dictionaries are for general purposes, we need to build our own dictionary for healthcare service domain.

We collect user reviews on medical services from NHS Choices for building our own dictionary (NHSdict). We collect randomly selected 2,163 reviews from the website. We assume negative reviews as with 1 or 2 stars and positive reviews as with 4 or 5 stars. We use 408 negative reviews and 408 positive reviews to measure the influence of a word on the classification of its review. We pre-process the reviews as we did previously. We use logistic, lasso, ridge and elastic regression [34] to make a model for classifying reviews into positive or negative. The independent variables of the regressions are terms from the reviews and the coefficient values are their sentiment scores.

We perform 10-fold cross validation for comparing the accuracy of the sentiment scores from the regressions. We simply summate scores of words which are contained in a review. Then we classify the review as negative if the summated score is less than zero, otherwise we classify it as a positive review. The classifications of the sentiment scores from the ridge regression outperform those from other regressions. Thus, we use the sentiment scores of words from the ridge regression for NHSdict.

Step 5: Sentiment Analysis
We measure sentiments of tweets by utilizing AFINN and NHSdict. We simply summate the sentiment scores of words in a tweet and then take the average of the sentiment scores according to the SERVQUAL constructs. Although AFINN has sentiment scores range between −5 to 5, the sentiment scores of NHSdict range between 0.1 to −0.1. To compare sentiment scores using two dictionaries, we standardize the sentiment scores by transforming the scores to z distribution. Then we merge AFINN and NHSdict to enlarge the size of dictionaries. The sentiment values of words in the merged dictionary are from the standardized values as described above. For terms in both dictionaries, we choose the sentiment value from NHSdict. We use all three dictionaries (AFINN, NHSdict, and the integrated dictionary) to compute the sentiments of the dimensions of the service quality.

5 Conclusion

In this paper, we proposed a systematic way of analyzing and tracking social perceptions of the quality of public services. Noisy social media data were filtered out by applying Doc2Vec and machine learning algorithms. The latent topics in social media data can be extracted and interpreted using the words belonging to the various topics. This paper provides a method for acquiring more reliable sets of topics and using the topics for classifying tweets into one of the SERVQUAL or other dimensions. We validated the performance of classifications and sentiment measuring using the training

data we obtained and the reviews from the NHS. Moreover, this paper provides an example of obtaining robust topic sets and using the topics to unravel the meaning of tweets.

For future research, it is imperative to investigate the results of applying more complicated methods, such as deep learning, to classify each tweet into the relevant construct. In this paper, we applied a term-matching method to calculate the sentiments and machine learning algorithms. Moreover, we have proven its applicability. In addition, sentence-based classification is more appropriate for longer expressions such as patient reviews. It will be interesting to apply this paper's method in that regard with some methodological alterations.

References

1. Swani, K., Brown, B.P., Milne, G.R.: Should tweets differ for B2B and B2C? An analysis of fortune 500 companies' Twitter communications. Ind. Mark. Manag. **43**(5), 873–881 (2014)
2. Chen, K., Kou, G., Shang, J., Chen, Y.: Visualizing market structure through online product reviews: integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. Electron. Commer. Res. Appl. **14**(1), 58–74 (2015)
3. Goh, K.Y., Heng, C.-S., Lin, Z.: Social media brand community and consumer behavior: quantifying the relative impact of user-and marketer-generated content. Inf. Syst. Res. **24**(1), 88–107 (2013)
4. De Silva, D.: Measuring patient experience. Health Foundation (2013)
5. Criado, J.I., Sandoval-Almazan, R., Gil-Garcia, J.R.: Government innovation through social media. Gov. Inf. Q. **30**(4), 319–326 (2013)
6. Greaves, F., et al.: Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual. Saf. **21** (7), 600–605 (2012)
7. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A conceptual model of service quality and its implications for future research. J. Mark. **49**(4), 41–50 (1985)
8. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. J. Retail. **64**(1), 12–40 (1988)
9. Teshnizi, S.H., Aghamolaei, T., Kahnouji, K., Teshnizi, S.M.H., Ghani, J.: Assessing quality of health services with the SERVQUAL model in Iran. A systematic review and meta-analysis. Int. J. Qual. Health Care **30**(2), 82–89 (2018)
10. Purcărea, V.L., Gheorghe, I.R., Petrescu, C.M.: The assessment of perceived service quality of public health care services in Romania using the SERVQUAL scale. Procedia Econ. Financ. **6**, 573–585 (2013)
11. Altuntas, S., Dereli, T., Yilmaz, M.K.: Multi-criteria decision making methods based weighted SERVQUAL scales to measure perceived service quality in hospitals: a case study from Turkey. Total Qual. Manag. Bus. Excell. **23**(11–12), 1379–1395 (2012)
12. Al-Borie, H.M., Sheikh Damanhouri, A.M.: Patients' satisfaction of service quality in Saudi hospitals: a SERVQUAL analysis. Int. J. Health Care Qual. Assur. **26**(1), 20–30 (2013)
13. Baeza-Yates, R., Ribiero-Neto, B.: Modern Information Retrieval: The Concepts and Technology Behind Search. Addison Wesley, Reading (2010)
14. Cao, Q., Duan, W., Gan, Q.: Exploring determinants of voting for the "helpfulness" of online user reviews: a text mining approach. Decis. Support Syst. **50**(2), 511–521 (2011)

15. Liu, Y., Bi, J.-W., Fan, Z.-P.: Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. Inf. Fus. **36**, 149–161 (2017)
16. Silge, J., Robinson, D.: Text Mining with R: A Tidy Approach. O'Reilly Media, Sebastopol (2017)
17. Cruz, R.A., Lee, H.J.: The effects of sentiment and readability on useful votes for customer reviews with count type review usefulness index. J. Intell. Inf. Syst. **22**(1), 43–61 (2016)
18. Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., Donaldson, L.: Harnessing the cloud of patient experience: using social media to detect poor quality healthcare: Table 1. BMJ Qual. Saf. **22**(3), 251–255 (2013)
19. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. **41**(6), 391–407 (1990)
20. Mehrotra, R., Sanner, S., Buntine, W., Xie, L.: Improving LDA topic models for microblogs via tweet pooling and automatic labeling. In: 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York (2013)
21. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012)
22. Bahja, M., Lycett, M.: Identifying patient experience from online resources via sentiment analysis and topic modelling. In: 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT). ACM, New York (2016)
23. Palese, B., Piccoli, G.: Online reviews as a measure of service quality. In: Pre-ICIS SIGDSAIFIP WG. Symposium, Dublin (2016)
24. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S.: Distributed representations of words and phrases and their compositionality. In: 26th International Conference on Neural Information Processing Systems (2013)
25. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. Presented at the Proceedings of the 31st International Conference on Machine Learning (ICML-14) (2014)
26. Niu, L., Dai, X.Y.: Topic2Vec: learning distributed representations of topics. http://arxiv.org/abs/1506.08422
27. Ramage, D., Rosen, E., Chuang, J., Manning, C.D., McFarland, D.A.: Topic modeling for the social sciences. In: 22nd International Conference on Neural Information Processing Systems (2009)
28. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**(Jan), 993–1022 (2003)
29. Lin, Z., Lu, R., Xiong, Y., Zhu, Y.: Learning ontology automatically using topic model. In: The International Conference on Biomedical Engineering and Biotechnology (2012)
30. Grün, B., Hornik, K.: topicmodels: an R package for fitting topic models. J. Stat. Softw. **40** (13), 1–30 (2011)
31. Nielsen, F.Å.: A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: The Proceedings of ESWC Workshop on Making Sense of Microposts (2011)
32. Liu, B.: Sentiment Analysis and Opinion Mining: Introduction and Survey. Morgan & Claypool Publishers, San Rafael (2012)
33. Mohammad, S., Turney, P.: Crowdsourcing a word-emotion association lexicon. Comput. Intell. **29**(3), 436–465 (2013)
34. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. STS, vol. 103. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-7138-7

Group Development Stages in Open Government Data Engagement Initiatives: A Comparative Case Studies Analysis

Arie Purwanto, Anneke Zuiderwijk, and Marijn Janssen(✉)

Delft University of Technology, Jaffalaan 5, 2628 BX Delft, Netherlands
{A.Purwanto, A.M.G.Zuiderwijk-vanEijk, M.F.W.H.A.Janssen}@tudelft.nl

Abstract. Citizens are increasingly using Open Government Data (OGD) and engaging with OGD by designing and developing applications. They often do so by collaborating in groups, for example through self-organized groups or government-induced open data engagement initiatives, such as hackathons. The successful use and engagement of OGD by groups of citizens can greatly contribute to the uptake and adoption of OGD in general. However, little is known regarding how groups of citizens develop in OGD engagement. This study aims at exploring and understanding the development stages of citizen groups in OGD engagement. To attain this objective, we conducted a comparative case study of group development stages in two different types of OGD engagement. Our cases show that leadership and diversity of capabilities significantly contribute to the success of citizen groups in OGD engagement. These findings suggest that connecting citizens having a diversity of expertise prior to the OGD engagement event helps to improve its effectiveness. This research is among the first to apply group development stages model in open data engagement studies and thus opening up new research opportunities concerning group developments in the open data literature.

Keywords: Open Government Data · Citizen engagement
Comparative case study · Group development · Self-organized
Government-induced · Hackathon

1 Introduction

Governments at different administrative levels (e.g., national, regional, local) are progressively opening up data to the public in the hope that citizens will use it [1]. Indeed, successful and sustainable use of Open Government Data (OGD) that contributes to solving societal problems hinges on citizens engagement [2]. We argue that citizen engagement is one move further than OGD use. Such engagement requires not only OGD use, e.g., locating, downloading, distilling, scrutinizing, and refining data [3], but also designing and developing OGD-based applications.

The development of applications by citizens based on OGD is often done by groups of people who collaborate [4]. Such groups can be self-organized, where the content and processes of engagement are determined by citizens who organize themselves and

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 48–59, 2018.
https://doi.org/10.1007/978-3-319-98690-6_5

engage in forms of collective action [5]. Examples of self-organized engagement include the Dutch's Open Spending [6] and the Indonesian's Kawal Pemilu [7] initiatives. In contrast, citizen engagement in OGD can also be government-induced. An example of such a government-induced initiative is a hackathon. In a hackathon initiative, governments determine when and where engagement takes place, and under which conditions citizens can engage [8].

In the open data literature, research in the socio-technical conditions of OGD utilization, both enabling and disabling factors, has widely been provided [1]. However, studies in the area of OGD engagement are lacking [1], especially regarding the development of groups of citizens who engage in these initiatives. Although individual citizens engaged in a group are motivated by different drivers [9], they strive to be successful in achieving the group's shared objectives. For example, in a hackathon, groups may not only attempt to win a competition and earn a prize but also indirectly contribute to solving problems raised by the hackathon organizers. Whereas in a self-organized OGD engagement initiative, groups may aim to contribute to solving a real-life problem they may face in daily life. However, there is scant knowledge of the group development and underlying factors that contribute to a group's success in the OGD engagement literature.

This study aims at exploring and understanding the development stages of citizen groups in OGD engagement. To attain this objective, we formulate the following research questions: "How do citizen groups develop in open government data engagement initiatives?" We conduct a comparative case study that involves two cases of OGD engagement in different settings. This study is among the first to apply group development stages model in comprehending citizen engagement in OGD initiatives. The results of our study advance the understanding of how policymakers should prepare and precondition the engagement initiative to stimulate more engagement groups.

2 Background

2.1 Open Government Data Engagement

Open data researchers usually define citizen engagement as open data use by citizens [10] that concerns various processes carried out to convert data to other outputs such as fact, information, data, interface, and service [11]. However, we argue that citizen engagement is one step ahead, involving not only OGD use, but also designing and developing OGD-based applications.

In public administration studies, researchers distinguish initiatives of citizen engagement with government policy between those that are self-organized and government-induced [5]. We argue that this distinction also applies to open data engagement because governments may operate using different models of data provision [8]. In the *government as a platform* model, the government limits its role only to the provider of open data infrastructure comprised of a web or portal offering access to data and tools for previewing, visualizing, or downloading data [8]. The government acts passively in this mode and presumes that others will use OGD, create applications on top of it and generate value [12]. This government mode seems to breed self-organized

OGD engagement initiatives. On the other hand, government-induced OGD engagement concerns the *government as open data activist* model in which the government not only provides the open data infrastructure but also promotes its use to citizens, the private sector, or the government itself [8]. In promoting open data use, such governments frequently organize supportive activities framed as a hackathon contest where citizens and businesses compete with each other to pitch an idea or the design of an application or an application prototype.

Self-organized Open Data Initiatives. Current open data literature is substantially lacking an overview of self-organized initiatives, and only little is known about this type of engagement. Self-organized engagement is somewhat a reaction to government-led processes or structures but utilizes the states' instruments (e.g., OGD portals and services) to attain citizens' objectives [5]. Organizing and sustaining such engagement requires the availability of two primary resources, time and money [13]. Therefore, only organized civil society that has access to sponsorships or donations can initiate and maintain self-organized engagement. Citizens initiated engagement such as Kawal Pemilu moved forward successfully because the initiators could radically minimize the costs incurred by using free open source software/platforms, utilizing social media platforms and applying crowdsourcing strategy [7, 14].

Government-Induced Open Data Events. This type of engagement typically takes form as hackathon events and aims to deliver economic value [15]. Since there is no agreement on the definition of an open data hackathon, we synthesize it based on selected literature [16–19] as follows. An open data hackathon refers to offline/face-to-face ideation competition sponsored by government agencies in a centralized location that brings together citizens with different backgrounds (e.g., programmers, designers, others) to intensively work collaboratively in small teams for a short amount of time (e.g., 12 h, 24 h, 2 days) to create artifacts (e.g., mockups, design, prototypes, applications) using OGD. Typically, at the end of the contest, each team presents/pitches the final idea in front of juries and sponsors, and a winning team earns a prize (e.g., money, investment).

In an open data hackathon, organizers and sponsors provide nearly all resources and support needed by the teams to work efficiently [16, 19, 20], including catering services, sleeping bags/area, comfortable facilities (gaming device, sports hall), internet connection, electricity (cables), and stationeries. Provision of technical support from open data providers or event organizers or sponsors is also common. These amenities are intentionally provided to support group development in the hackathon event.

2.2 Group Development Stages

Either in self-organized initiatives or hackathon events, the development processes of a citizen group/team would determine how they conceptualize a problem, brainstorm potential solutions, develop the preferred solution collaboratively and ultimately deliver it at the end. Self-organized initiatives might produce a ready-to-use application for society, whereas hackathons might offer various outputs based on the event's objectives (e.g., mockup, design, prototype, application, visualization). While current literature does not signal cue for group development in self-organized initiatives, on the

contrary, a small number of hackathon studies has started discussing the theme [21, 22]. However, both works do not specifically focus on how teams progress throughout the hackathon.

Studies on group development incorporate the investigation of group activities and how these activities evolve over the life of a group [23]. Stages or phases of group development are defined as the categorization of "the periods of time during which an identifiable set of activities occurs" [23, p. 122]. Although numerous models of group development have been proposed, Tuckman's [24] classical sequential stages is one of the most influential models recognized in the human resource development studies [25]. In this model, Tuckman [24], focusing on interpersonal relationships and task activity, postulated a four stage of group development namely *forming*, *storming*, *norming*, and *performing* (see Fig. 1). Tuckman [24] further posited that effective group functioning requires successful formation of each stage and transformation from one stage to another.

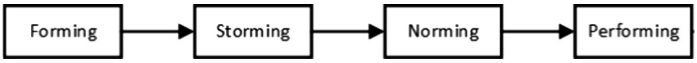

Fig. 1. Tuckman [24] model (adapted from Bonebright [25])

Forming. Tuckman [24] described the first stage as *testing* and *dependence* of interpersonal relationships (group structure) among group members and *orientation* to the task activity. Group members attempt to discover acceptable behaviors based on the reactions of the group leader and other members. Once the boundaries are discovered, a member becomes dependent on the guidance and support from the leader(s) and pre-existing norms. Group members attempt to identify relevant tasks and ways to accomplish the tasks by determining information required to deal with the tasks and how the information can be acquired.

Storming. The second stage is characterized by *intragroup conflict* related to group structure and *emotional response* to task demands that lead to the *lack of unity*. Group members express their individuality and oppose the formation of group structure by becoming resistant toward one another and group leader(s). The discrepancy between individual's interest and orientation demanded by the tasks leads to emotional reactions and resistance to the tasks. However, Tuckman [24] considers that this stage would be less visible in groups working on intellectual tasks.

Norming. The third group structure stage is identified as the development of *group cohesion,* and the task activity development is characterized as the *open exchange* of relevant interpretations. A member accepts group structure and the individuality of fellow members. New group-generated norms endorsing *harmony* to ensure the group's existence are the results of the acceptance. Group members are open to discussing themselves and others' and their opinions to generate an alternative interpretation of tasks.

Performing. In the fourth stage, the development of group structure is labelled as *functional role-relatedness,* and the development of task-activity is identified as the *emergence of solutions*. Members adopt and play roles after learning from one another socially in the preceding stage. Role structure becomes an instrument that can direct the group as a problem-solving entity. Constructive actions that lead to successful tasks accomplishment (solutions) are seen in this stage.

3 Research Methodology

3.1 Case Study Design

The research aims at understanding and exploring the citizen's group developmental stages which are presently little understood in contemporary open data engagement context. As a result, the aim might be attained using qualitative approaches and cannot be achieved using quantitative inquiries such as a survey. Although the study was informed by a prior model of group development stages [24], it is unclear whether the seminal model applies to different types of OGD engagement. Therefore, the case selection aimed at finding cases that concern the citizen's group developed in OGD engagement initiatives and providing variation in contextual factors (self-organized and government-induced) that enable polar cases. Case studies are appropriate for research trying to answer "how" or "why" questions about contemporary events over which the researcher has little or no control [26].

We selected cases that concern OGD and groups of citizens engaging in the OGD initiatives. The cases must involve groups representing different types of OGD engagement. The cases should also include groups that accomplish a set of contextual objectives. To enable comparison and contrast between cases, we selected two cases that are varying contextually: the Kawal Pemilu group that exemplifies the self-organized engagement and the PacMan team that epitomizes the OGD government-induced engagement (hackathon). Table 1 provides a brief overview of the selected cases.

The first case involves a group of citizens who voluntarily developed an OGD-based application and used it to digitize the results of Indonesian's 2014 presidential election. The group comprises two teams, a developer team of five technologists who built the application and a volunteer team of 700 persons who used the application. The successful digitization of election results, covering 97.91% of 478,829 votes, in only six days made Kawal Pemilu a prominent example of citizen engagement [7].

The second case concerns a team of citizens who participated in a Dutch's open education data hackathon, Hack de Valse Start, held on 3 March 2018 for twelve hours (from 8 AM until 8 PM). PacMan comprises five persons with diverse backgrounds and capabilities who worked in a collocated room of a high school building situated in the outskirt of the Amsterdam city. The group, competing with six other teams, won the second prize for visualizing averaged national exam scores data against averaged teacher advice data at the school level and providing an analysis of the visualization.

Table 1. A brief overview of cases.

Characteristics	Kawal Pemilu	PacMan
Engagement type	Self-organized	Government-induced (hackathon)
Country location	Indonesia	The Netherlands
Number of citizens	705	5
Proximity	Virtual (geographically dispersed)	Face-to-face (collocated)
Period	14 days (9–22 July 2014)	12 h (3 March 2018)
Background	Indonesian	Dutch, Russian, Romanian, and Indonesian
Output	OGD-based application	OGD visualization and analysis
Objectives	Digitizing all election results	Winning the competition
Open data domain	Election results provided by the Election Commission of Indonesia	Education data provided by the Dutch Education Inspectorate and Central Bureau of Statistics

3.2 Data Collection and Analysis

We collected various types of qualitative data from multiple sources of evidence at several points in time, to enhance construct validity as much as possible [26]. In both groups, the first author conducted participant-observations by actually participating in the engagement: as a volunteer in the Kawal Pemilu group and as a member of the PacMan team. Gaining actual access to these teams provide a distinctive opportunity to understand the group development from the perspective of an *insider* since post-factum comprehension of interpersonal relationships and task activities is non-trivial [26]. The researcher used online observation through the Facebook (FB) platform because the Kawal Pemilu group was developed entirely using the platform.

Table 2 provides an overview of the case information sources, including documents, interviews, participant observations and tangible artifacts. Fifteen semi-structured interviews were conducted with the Kawal Pemilu group members from October 2017 until February 2018. All interview sessions were recorded as agreed by the interviewees and transcribed. The author also conducted four unstructured, informal interviews with the PacMan team members during the hackathon. Since the data collected include personal data from both groups concerning privacy and confidentiality, the first author was only approved by group leaders to disclose their data.

We divided our analysis into two phases. First, we analyze the data using provisional manual coding to capture the development process of both groups based on Tuckman's [24] stages. Finally, we categorized the codes into two groups: (1) interpersonal relationships and (2) task activities associated with the developmental stages as indicated by the model.

Table 2. Data collection strategy.

Data Source	Kawal Pemilu	PacMan
Documents	Email correspondence, records from question-and-answer site, news, personal blogs, documents from the online collaboration platform	Notes, news, documents from the online collaboration platform
Interviews	Fifteen individuals (two developers and thirteen volunteers) semi-structured interviews	Four individuals (one leader, one data scientist, one translator, one supporting) unstructured, informal interviews
Participant-observations	FB secret group	Twelve hours face-to-face meeting
Tangible artifacts	Election outcome website	Presentation file

4 Results

4.1 The Development of the Kawal Pemilu Group

The Kawal Pemilu group was invented by Ainun; an Indonesian-national data scientist lived in Singapore on 9 July 2014 immediately few hours after competing presidential candidates declared their victories. Ainun recruited four Indonesian developers living in different countries (i.e., Australia, the Netherlands, and United States) to build the digitization application and 700 Indonesian volunteers around the world to digitize election results using the application. Since two teams were involved in the Kawal Pemilu, we presented the results as separated but connected processes of both teams (see Table 3).

Table 3. The Kawal Pemilu group development.

Stages	Developer team		Volunteer team	
	Interpersonal relationships	Tasks activity	Interpersonal relationships	Tasks activity
Forming	Members were friends who trust each other's integrity	The norm was the due date of the official election victor announcement	Members were mostly unknown to others	The norm was the due date of the official election victor announcement
Storming			Conflicts arose among members related to a political stance	Members resisted imbalance tasks and preferred to do only interesting tasks
Norming	Discussions about the application requirement decision were organized	Due date was relaxed, and new technological decisions were made	New role, verifier, was established to examine volunteers' works	Pre-existing norms were still in place
Performing	Roles were established and adopted	Efforts based on roles were taken to develop, sustain and maintain the application until most ballots were digitized	Two roles (inputter and verifier) were established and adopted	Volunteers strived to digitize all ballots and verifiers validate the digitization results and report errors

Two of the developers were Ainun's close friends, and both trusted Ainun's integrity. The other two developers were invited by one of Ainun's friends. Social relationships were well developed among members of this group. Contrary, a volunteer might know several other volunteers but rarely knew all of them due to a large number of persons involved. No guidance was determined other than the due date, 22 July 2014, set by the Election Commission of Indonesia to officially announce the election victors.

Conflicts among volunteers arose as a form of distrust towards each other's political stance and interests. Some volunteers, siding with one of the election candidates, suspected that other volunteers, supporters of the opposed candidate, would damage the digitization initiative by deliberately inputting an incorrect number of ballots. Volunteers resisted the tasks distribution due to two issues as follows. First, some volunteers prioritized inputting the results from the regional area where they or their families or friends were living in. Second, the number of voting booth varied across regional areas and might lead to imbalance tasks distribution. The densely populated area was likely to have more booths and thus more ballots to be digitized.

On 9 July 2014, the developer team started brainstorming and discussing the idea and design of the application, using an online collaboration tool. An external expert was invited to the discussion sessions. The discussions occurred until 14 July 2014 and were entirely positive and technical towards choosing the right algorithms for verifying errors, incentivizing volunteers, and preventing incorrect data service invocation. Although at some point members disagreed with other's opinions, the disagreements were seen as intellectual dialogs, not interpersonal conflicts. The pre-existing norms evolved into new norms as a result of the discussions: the due date was relaxed, and new technological decisions were made. New role, verifier, was established and followed up by recruitment among volunteers. Verifiers were grouped into small teams and tasked to examine input made by volunteers and correct erroneous inputs. A verifier team's results were further re-examined by another team to improve data reliability.

Team (developer, volunteer, and verifier) members quickly understood their respective roles and performed tasks accordingly. All efforts were made to sustain the Kawal Pemilu's website until the digitization of election results finished on 18 July 2014.

4.2 The Development of the PacMan Team

The PacMan team was initiated by Johannes, an educational journalist working for De Correspondent, a Dutch news website. Johannes randomly asked nearby participants to join his team and further asked interested participants to get to know each other's strength by explaining their background and specialization. Four participants including the first author agreed to form a team with Johannes. The first member was a data scientist from Russia, working for a Dutch travel aggregator company, who has participated in numerous hackathon events. Another member was a Dutch and an employee of a municipality in the Netherlands who worked in the education field. The third member was a workshop organizer from Romania, working for promoting open data use through "maker" arts. Table 4 describes the development of the PacMan team.

Johannes, henceforth the team leader, initiated the brainstorming of interesting societal problems that can be explored and exploited as the team's final product. Although three members were not Dutch persons, they contributed to the discussions. The data scientist viewed the topic proposal from technical viewpoints and sometimes disagreed with the leader since the topics were not supported by available data. The first author clarified the current government's educational policy and the data visualization that will be pitched. The municipality employee added several local social issues to consider in the visualization. At the end of the discussion sessions, new team norms were added: a visualization to compare national exam scores against school advice and to provide a preliminary indication of the causes of deviation between scores and advice.

Table 4. The PacMan team development.

Stages	Interpersonal relationships	Tasks activity
Forming	An initiator, henceforth the leader, formed the team through an introduction session among interested candidates	Members tried to understand the team goal informed by the scheduled pitch session at the due time
Norming	Members expressed their opinions in intellective discussions guided by the leader	New norms were agreed
Performing	Members understood their and others' respective roles	Members acted according to their roles to meet the deadline

Roles were understood and performed accordingly. Johannes searched for relevant data and handed over the first author. The first author examined the data and supplied relevant data (e.g., statistical socioeconomic data) to the data scientist who coded the visualization. The municipality employee helped translate the metadata written in the Dutch language to English and explain the meaning to the data scientist. The workshop organizer prepared online collaboration tools and design the presentation for the pitch session. Fifteen minutes before the pitch started, the final presentation file was completed and submitted to the hackathon organizers. The leader delivered the final presentation in Dutch to provide contextual meaning to the team's output.

5 Discussion and Conclusion

As indicated by Tuckman [24], the Kawal Pemilu developer and PacMan teams, working on intellective tasks namely developing an application and designing a visualization, progressed through forming, norming and performing stages. In contrast, the Kawal Pemilu volunteer teams evolved around four stages including the storming stage. Digitizing election results seemed to be personal tasks because some volunteers preferred to digitize specific regions and tended to take sides in the election. Despite

different time duration of the engagements under study, these results signal the relevance of the stages in both virtual and face-to-face groups development.

Different interaction factors appeared in both cases. While the impact of duration on these interactions needs to be studied further, the verbal and tangible presence of nonverbal cues in the PacMan team seemed to enhance the communication among its members. In addition to the conflicts of personal interests seen in the storming stage, communicating virtually with strangers could hamper the interactions of the Kawal Pemilu group members. The use of *emoticons* in FB platform might help improve participants' perceptions towards others' emotion, attitude, and attention intention [27], thus decreasing the communication barrier. Nevertheless, further studies are needed to test these propositions since literature suggest that computer-mediated interaction lacks cues to reduce the communication perception problems [27].

We propose three non-exhaustive underlying factors that appear to contribute to the success of group development stages. *Leadership* roles, naturally played by Ainun and Johannes who actively sought for personnel that might help them achieve the group's objectives, contribute to quicken the group formation. *Beforehand interaction* of participants may help reduce communication issues in forming the group and identify roles needed to perform tasks. *Diverse capabilities*, technical (e.g., programming) and domain-related (e.g., election, education systems) skills and knowledge, enhance progress in tasks performance of OGD engagement and its context.

Policymakers should consider the above factors in promoting OGD engagement. Although locating a leader is non-trivial since open data users are commonly unknown to policymakers [10], surveying open data communities may lead to potential champions and enable informing them early about OGD initiatives. Providing an online platform that connects open data user groups and enables them to interact with each other may facilitate interactions and help them know other's profiles before OGD engagement is actually carried out.

In addition to the discussion above, we are aware of the limitations of this study concerning the use of a participant-observation strategy. In the Kawal Pemilu case, the first author was able to be an external observer since nearly all activities were performed virtually and involved a large number of participants. Contrary, the author's participation in the PacMan case might lead to advocacy roles that contradict the practice of good social science [26]. However, the researcher was able to play observer role until the performing stage that requires more technical activities than social relationships.

This study is an initial step in understanding how citizens engage in OGD initiatives from a linear group development stages perspective. Although Gersick [28] proposes a more complex rivalry model, incorporating temporal aspects of group development, she confirms its similarity with Tuckman's [24] stages. Therefore, future research should consider Gersick's [28] model on the transitions of citizen OGD engagement groups. Another path for future research concerns the inclusion of more, similar and polar, cases to test theories or models and the validation of the developmental stages model using quantitative approach.

References

1. Safarov, I., Meijer, A., Grimmelikhuijsen, S.: Utilization of open government data: a systematic literature review of types, conditions, effects and users. Inf. Polity **22**, 1–24 (2017)
2. Dietrich, D.: The role of civic tech communities in PSI reuse and open data policies (2015)
3. Cornford, J., Wilson, R., Baines, S., Richardson, R.: Local governance in the new information ecology: the challenge of building interpretative communities. Public Money Manag. **33**, 201–207 (2013)
4. Susha, I., Janssen, M., Verhulst, S.: Data collaboratives as "bazaars"? A review of coordination problems and mechanisms to match demand for data with supply. Transform. Gov.: People Process Policy **11**, 157–172 (2017)
5. Edelenbos, J., van Meerkerk, I., Schenk, T.: The evolution of community self-organization in interaction with government institutions: cross-case insights from three countries. Am. Rev. Public Adm. **48**, 52–66 (2018)
6. Hartog, M., Mulder, B.: Accountable and responsible disclosure of financial open government data: open spending initiatives enhancing civic engagement. In: Kommers, P., Peng, G.C. (eds.) International Conference ICT, Society and Human Beings 2017, pp. 127–133. IADIS, Lisbon (2017)
7. Brajawidagda, U., Chatfield, A.T.: Roles of social media in open data environments: a case study of the 2014 Indonesian presidential election voting results. In: 25th Australasian Conference on Information Systems, ACIS, Auckland, New Zealand (2014)
8. Sieber, R.E., Johnson, P.A.: Civic open data at a crossroads: dominant models and current challenges. Gov. Inf. Q. **32**, 308–315 (2015)
9. Juell-Skielse, G., Hjalmarsson, A., Johannesson, P., Rudmark, D.: Is the public motivated to engage in open data innovation? In: Janssen, M., Scholl, H.J., Wimmer, M.A., Bannister, F. (eds.) EGOV 2014. LNCS, vol. 8653, pp. 277–288. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44426-9_23
10. Susha, I., Grönlund, Å., Janssen, M.: Organizational measures to stimulate user engagement with open data. Transform. Gov.: People Process Policy **9**, 181–206 (2015)
11. Davies, T.: Open data, democracy and public sector reform: a look at open government data use from data.gov.uk. Open Data Impacts - Research blog of Tim Davies (2010)
12. Linders, D.: From e-government to we-government: defining a typology for citizen coproduction in the age of social media. Gov. Inf. Q. **29**, 446–454 (2012)
13. Brady, H.E., Verba, S., Schlozman, K.L.: Beyond SES: a resource model of political participation. Am. Polit. Sci. Rev. **89**, 271–294 (1995)
14. Purwanto, A., Janssen, M., Zuiderwijk, A.: Towards an open government data success model: a case study from Indonesia. In: 17th European Conference on Digital Government, Lisbon, Portugal, pp. 154–162 (2017)
15. Johnson, P., Robinson, P.: Civic hackathons: innovation, procurement, or civic engagement? Rev. Policy Res. **31**, 349–357 (2014)
16. Concilio, G., Molinari, F., Morelli, N.: Empowering citizens with open data by urban hackathons. In: 7th International Conference for E-Democracy and Open Government (CeDEM), Krems, Austria, pp. 125–134. IEEE (2017)
17. Hartmann, S., Mainka, A., Stock, W.G.: Opportunities and challenges for civic engagement: a global investigation of innovation competitions. Int. J. Knowl. Soc. Res. **7**, 1–15 (2016)
18. Lara, M., Lockwood, K.: Hackathons as community-based learning: a case study. TechTrends **60**, 486–495 (2016)

19. Soltani, P.M., Pessi, K., Ahlin, K., Wernered, I.: Hackathon – a method for digital innovative success: a comparative descriptive study. In: Devos, J., Haes, S.D. (eds.) 8th European Conference on Information Management and Evaluation (ECIME). Academic Conferences and Publishing International Limited, Ghent (2014)
20. Briscoe, G., Mulligan, C.: Digital Innovation: The Hackathon Phenomenon. Queen Mary University of London (2014)
21. Richterich, A.: Hacking events: project development practices and technology use at hackathons. Converg.: Int. J. Res. New Media Technol. (2017). http://journals.sagepub.com/doi/full/10.1177/1354856517709405
22. Trainer, E.H., Kalyanasundaram, A., Chaihirunkarn, C., Herbsleb, J.D.: How to hackathon: socio-technical tradeoffs in brief, intensive collocation. In: 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, California, USA, pp. 1118–1130. ACM (2016)
23. Miller, D.L.: The stages of group development: a retrospective study of dynamic team processes. Can. J. Adm. Sci. **20**, 121–134 (2003)
24. Tuckman, B.W.: developmental sequence in small groups. Psychol. Bull. **63**, 384–399 (1965)
25. Bonebright, D.A.: 40 years of storming: a historical review of Tuckman's model of small group development. Hum. Resour. Dev. Int. **13**, 111–120 (2010)
26. Yin, R.K.: Case Study Research: Design and Methods. SAGE Publications Inc., California (2014)
27. Lo, S.-K.: The nonverbal communication functions of emoticons in computer-mediated communication. CyberPsychol. Behav. **11**, 595–597 (2008)
28. Gersick, C.J.G.: Time and transition in work teams: toward a new model of group development. Acad. Manag. J. **31**, 9–41 (1988)

Managing Standardization in eGovernment: A Coordination Theory based Analysis Framework

Dian Balta[1(✉)] and Helmut Krcmar[2(✉)]

[1] fortiss GmbH, Guerickestr. 25, 80805 Munich, Germany
balta@fortiss.org
[2] Informatics 17 - Chair for Information Systems,
Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany
krcmar@in.tum.de

Abstract. Standardization plays an important role for a smooth organizational, semantic, technical and legal interoperation of eGovernment services. Still, standardization struggles with the complexity of administrative procedures that have to be supported and provided online. Managing this complexity poses a crucial challenge for an efficient and effective eGovernment. Different management frameworks have been developed, but the progress of standardization in practice is still perceived as insufficient. To address the challenge of standardization management in eGovernment, we propose a coordination theory based framework for analysis. It consists of three coordination modes distinguished based on their mechanisms and relevant context dimensions. In our research approach, we interpret a single case study of a standardization project in Germany where artefacts were developed from a pragmatism perspective. We discuss our findings and conclude on implications for research and practice.

Keywords: Standardization · Coordination · Management

1 Introduction

Standards play an important role in effective and efficient service provision in government [1–3]. For instance, standards are applied to harmonize business processes [4] and to efficiently share information [5] at the right time with the right people and in the right quality [6] by enabling the interconnectivity of the information systems that underlie the provision of services to citizens and companies [7]. Choosing a suitable standardization approach and coordination of the corresponding process [8, 9] based on the context and expected benefits of an eGovernment project [10, 11] are a prerequisite for its success [1, 12, 13].

Coordination of standardization faces a number of challenges in the government domain due to the adherent complexity of administrative procedures [2, 14, 15]. Research on standardization in organizations [16] and information systems management [17] has emphasized measures to counter complexity challenges (e.g. [12, 18]), but the measures' applicability in a government context is limited by differences between private organizations and government bodies [19, 20]. In particular, little is

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 60–72, 2018.
https://doi.org/10.1007/978-3-319-98690-6_6

known of what should be coordinated in order to manage the dependencies between actors involved in standardization of administrative procedures. Extant research addresses predominantly how and why to coordinate [21], e.g. how to coordinate applying particular information system architectures (e.g. [22–24]), how coordination evolves in public administrations (e.g. [25]) and why coordination is crucial for an eGovernment driven transformation of administrative procedures (e.g. [26]).

Our research goal is to develop a coordination theory based framework for analysis and effective standardization of administrative procedures in eGovernment that comprises three modes of coordination. Each mode is described by a set of applicable mechanisms and is distinguished from other modes based on a set of dimensions that allow for justifying its application in a given situation. To develop the framework, we conduct a qualitative study on coordination by interpreting artefacts developed in a standardization project. To describe the applicability of the framework, we exemplify the insights gained in the project under study.

This paper is structured as follows. In the next chapter we describe the theoretical background of the paper. Next, we describe our research approach. In chapter four, we derive our coordination theory based framework. Next, we present an exemplary application of the framework. Finally, we discuss our results and conclude on implications for theory and practice.

2 Theoretical Background

2.1 Standardization

We define a standard as "a uniform set of measures, agreements, conditions, or specifications between parties" [27]. The process of reaching a standard encompasses stabilizing and solidifying its definition and boundaries [12, 28], i.e. standardization represents "the activity of establishing and recording a limited set of solutions to actual or potential matching problems directed at benefits for the party or parties involved balancing their needs and intending and expecting that these solutions will be repeatedly or continuously used during a certain period by a substantial number of the parties for whom they are meant" [29].

Standardization can be divided into five general levels of solutions for the parties involved [30, 31]. First, reference models are defined on the first level, followed by a set of functions and technologies on the second level. The third level is represented by a functional profile that determines the user group's requirements. The fourth level is described by a system profile that determines the requirements of specific classes of users. Finally, a specific application implementation is represents the fifth level of standardization.

Since standardization involves information systems, process and organizations and different federal layers [2, 32, 33], standards at different levels as solutions for managing information and underlying systems are a necessity for improving the maturity of digitalized government services [34]. Although this necessity represents one of the drivers for an increased attention to standardization in government [8, 35], standardization in eGovernment practice is challenging. In particular, standardization is

challenged by the complexity of government procedures [2, 14] and is often ineffective in practice [36, 37]. Various approaches to information management that are tailored to the context of government bodies through standardization have been developed in response [38, 39]. For example, frameworks like SAGA in Germany [40], with its focus on software specifications and development methods in the context of government bodies, and the European Interoperability Framework [41] have been proposed to management practice. Consequently, addressing standardization at different federal layers and at different levels would benefit from an effective coordination.

2.2 Coordination

Coordination is seen as a central concept regarding the design of organizations (e.g., [42–44]) and work (e.g. [45]). It represents a topic of study at different levels, e.g. individual, community, and in multiple disciplines [46], including IS, management and e-government. Prior research on information systems management and organization design has focused primarily on assessing the efficacy of various coordination mechanisms [47] as well as on the causality of their outcomes [44]. For instance, it has addressed integrative conditions of coordination such as accountability, predictability, and common understanding [45], the applicability of different coordination modes in the context of knowledge-intensive work [48–50], mechanisms targeting IT use and capabilities that are based on steering committees and communication policies [51, 52] as well as sharing of knowledge and spanning of systemic and organizational boundaries [53–55].

With regard to extant literature and the numerous perspectives on coordination [44–46, 56], this paper addresses *coordination* as a means for "*managing dependencies between activities*" [46]. Since coordination takes place between individuals [49, 57], we use the term *coordination of administrative procedures* to denote the management of dependencies between activities involving persons and the resources (e.g. information, knowledge) they exchange in regarding an eGovernment administrative procedure. Further, coordination is exerted through different mechanisms that can be predefined or emerge based on the context there are applied in [57–59].

3 Research Approach

The research described in this paper is conducted as a qualitative study of coordination in information systems [60, 61], epistemologically rooted in two paradigms: interpretation [62, 63] as well as pragmatism [64, 65]. We choose this approach in accordance with [65, 66] for two reasons. On the one hand, the authors were involved in dialogical action during the project Federal Information Management (FIM) under study in their role as scientific consultants [67]. This allowed for an in-depth study of a single case. On the other hand, we aim at the development of a theory based artifact combining interpretation of work [63] well as practical inquiry [65].

3.1 Case Background

In Germany, administrative procedures are situated on federal, state and community level and, with regard to administrative traditions [33], can be described as quiet heterogeneous between the levels. They often differ in their design and execution in public administrations situated at the same level, e.g. two communities in the same state could exchange information with involved actors during a procedure in a totally different way depending on public administration size or the information technology landscape of the community.

In an effort to manage the complexity of numerous and heterogeneous administrative procedures, the federal and state public administrations in Germany started in 2012 the project FIM to standardize procedures at all federal layers on a reference model level and is currently applied in the design of an interoperable government service platform. Each administrative procedure was described as a combination of a service, an underlying process to execute it and a set of forms for input or output data in the process. Since the project had the management attention at chief information officers' level of the German Federation as well as of all German States, we were able to interact intensively with experts and practitioners at all federal levels.

3.2 Data Collection and Analysis

Empirical data collection and analysis took place in several phases (cf. Fig. 1). First, potentially relevant standardization initiatives in Germany were identified (phase I). We conducted expert interviews during project meetings and eGovernment practitioner conferences in Germany as well as an analysis of documents regarding standardization initiatives presented in German public meetings, conferences and congresses. Based on our findings, we identified addressees for a preliminary survey of relevant initiatives. The survey was conducted by e-mail or phone with employees in IT management positions at federal or state public administrations and resulted in 40 potentially relevant initiatives[1].

In a second step, data was obtained through the analysis of websites and publications regarding each initiative as well as by interviewing public administration employees based on the survey (phase II). Prior to each interview, we sent a scheme for analysis of components and characteristics of standardization aimed at by each initiative that we developed. Consequently, we were able to conduct a total of 40 semi-structured in-depth interviews in person or per telephone, with duration between 15 and 45 min per interview. The interview data was used to complement the information on the initiatives obtained during the first step. Eventually, we identified 30 relevant initiatives in German public administration and to derive our initial framework.

Next, we developed coordination components and standardization characteristics as a framework for developing reference models for administrative procedures, documented it and discussed its application with experts in workshops, focus groups or individual interviews (phase III). Based on this framework, we created text documents, presentations as well as workshop protocols in dialogue and action together with

[1] cf. http://fim.yourresearchportal.com/ for the initial data set applied in our research.

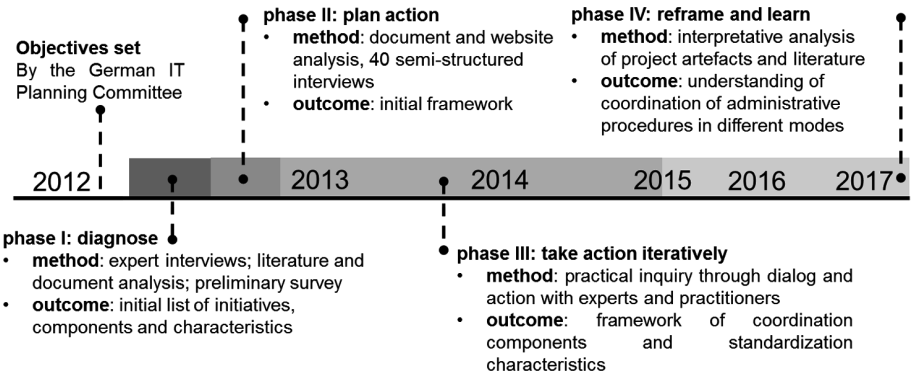

Fig. 1. Research phases of data collection and analysis

experts and practitioners [67] regarding the application and evaluation of our framework. This "practical inquiry" [65] allowed us to mix and improve both practitioner's and researcher's experience on coordination of administrative procedures through the mechanism of standardization. Consequently, we were able to provide for a satisfactory remedy to a real problem [67] and the framework as an artefact represented "constructive knowledge" [65] towards an effective coordination of complex administrative procedures.

Finally, we studied all project artefacts and literature in an interpretative phase (phase IV) through an argumentative-deductive analysis [68]. Thus, we incrementally adjusted our interpretation [63, 69] by reframing our artefacts taking a coordination theory perspective. Thus, we were able to learn and enhance our understanding [65] and to contribute to theory towards effective coordination through standardization of complex administrative procedures considering different coordination modes.

4 A Coordination Theory Based Framework for Standardization Management

Three *modes of coordination* can be distinguished [48, 49, 70]: two modes of explicit coordination referred to as *planning* and *feedback* as well as a third mode, complementary to the previous two modes, referred to as *tacit coordination* [48, 49] (cf. Table 1). The first mode, explicit coordination by planning, involves a designer, who partitions activities into capsuled modules and builds interfaces between the modules [48, 49, 71]. Applicable planning mechanisms involve pre-established plans, schedules, forecasts, formalized rules, policies etc. For example, consider the electronic data interchange protocol as a mechanism of coordination by planning. It involves a number of actors (companies with their employees) that accomplish tasks in their own modules and exchange information based on the protocol as an interface between the modules.

The second mode, explicit coordination by feedback, represents "mutual adjustments" between the individuals that take place "upon new information" [43, 72]. Applicable mechanisms in this mode involve ongoing communication: an on-demand

Table 1. A coordination theory based framework for standardization management

	Planning mode (explicit)	Feedback mode (explicit)	Knowledge exchange mode (tacit)
General mechanism	Partitioning of tasks into modules and defining interfaces	Establishment of an on-demand, face-to-face (or virtual) ongoing communication	Exchange and adaption of boundary objects, transfer of resources
Common ground between actors	Minimal, constant level	Constant updates	Knowledge transfer and updates by adaptation to context
Communication	Impersonal, "one-time" communication, modules and interfaces	Personal or in group ongoing communication	Different possible ways to communicate, but explicit communication is not a prerequisite
Environment	Stable, pre-established tasks and responsibilities	Dynamic, innovative tasks under volatile conditions	Dynamic, innovative tasks under volatile conditions
Interdependency type	Pooled or sequential interdependency	Reciprocal interdependency	Reciprocal or team interdependency
Example	Public administrations in the EU using the eIDAS message format	A telephone call or a meeting, an online video conference	A software codebase [50], team staffing with an experienced member from another team [48]

face-to-face (or virtual) personal or group communication takes place. For example, a scheduled or unscheduled group meeting takes place in order to cope with challenges resulting from new information relevant for the task at hand. The actors that take part in the meeting adjust their knowledge by active communication.

The third mode, *tacit coordination*, involves the notion that coordination takes place in a tacit manner. Mechanisms in this mode build or leverage shared (tacit or explicit) knowledge [48, 49, 73]. Sharing knowledge is accomplished through boundary objects [45, 49, 74] and can span across organizational [53–55]. For example, a team can be staffed with an experienced member from another team who is able to transfer context based knowledge into the task at hand in the new team. An experienced team member can introduce a boundary object such as a software codebase [50] and other team members can coordinate by adapting their own tasks according to the codebase without the need of a particular coding protocol or an intense ongoing communication. Consequently, tacit coordination takes place as an indirect communication between the actors while the task at hand is being implemented [49]. Moreover, it helps to share task specific knowledge between contexts [48].

For each of the three modes, the management of dependencies between activities involves an analysis of the situation of coordination [46]. The situation comprises an

understanding of which activities and dependencies should be coordinated [46]. It can be described along the following dimensions: (i) the *common ground* of the actors involved, (ii) the applicable means of *communication*, (iii) the aspects of the *environment* where the coordination takes place as well as (iv) the type of interdependencies to be considered [46, 75]. We refer to the dimensions of the situation of coordination to emphasize the differences between the modes as follows (cf. Table 1).

The first dimension common ground is denoted as knowledge that is or will be shared in order to allow for reciprocal predictability of action [49, 73, 76]. Coordination by planning involves a minimal level of common ground that is constant, i.e. the actors share knowledge to a limited extent and rarely update it (e.g. an information exchange protocol). In contrast, feedback and tacit coordination involve constantly updates of the knowledge shared by individuals: feedback, trough ongoing communication, and, regarding tacit coordination, through knowledge adaptation to context.

The second dimension – communication - considers how the actors involved exchange information upon coordinating task accomplishment [42, 77]. Coordination by planning and feedback represent two poles regarding this dimension. Since planning involves upfront definition of modules and interfaces (as a means to exchange resources such as knowledge between the modules), the communication is impersonal and takes place on an irregular pace (e.g. at the beginning of the task accomplishment). On the contrary, coordination by feedback comprises ongoing personal or group communication with no pre-established protocols to follow. Similar to feedback, tacit coordination might involve personal communication or any pre-established protocols. Still in this mode, communication is not ongoing or even not a prerequisite, since the common ground can be built without it through knowledge adaptation to context (e.g. shared source code base).

Environment, the third dimension of a situation of coordination, considers the contextual setting in which the tasks take place. Thus, it includes external factors that influence the actors and their activities. Basically, an environment can be dynamic or stable, i.e. (i) it accounts for innovative tasks that require novel common ground and do not rest upon pre-established plans or responsibilities or (ii) it accounts for common well-known tasks that can be accomplished based on pre-defined modules and interfaces. Feedback and tacit coordination account for mechanisms applicable to a more dynamic environment, while mechanisms of the planning mode suit a more stable environment.

The fourth dimension – interdependency type – addresses how each actor depends on the other actors during task accomplishment. Four dependency types can be distinguished: pooled, sequential, reciprocal and team interdependency [75, 78, 79]. Coordination by planning is suitable to interdependencies of the first and second type, i.e. one central actor (acting as a pool) coordinates the actions of other actors (according to pre-defined modules) or actors accomplish task in a sequence with interfaces between each actor (e.g. according to a predefine protocol). Feedback is suitable to the coordination of reciprocal dependencies between actors, i.e. exchange of resources relevant for task accomplishment takes place between all actors (e.g. a group meeting). Similar to feedback, tacit coordination is suitable to reciprocal dependencies. Moreover, it is suitable to team interdependency, i.e. reciprocal differences exist and actors simultaneously collaborate towards task accomplishment.

Since tacit coordination is complementary to the both explicit coordination modes, in a particular situation of coordination a combination of the modes should be expected. Consequently, mechanisms applied in each mode should address the complementary function of the modes.

5 Exemplary Application of the Framework

We analyze the project FIM based on our framework (cf. Table 2) and emphasize a set of artefacts– targeted at an efficient and effective standardization using reference models of services, processes and forms–that resulted in the project. Hence, we aim at justifying each artefact according to its coverage of a particular mode of coordination.

Table 2. Analysis of the project FIM using our coordination theory based framework

	Planning mode (explicit)	Feedback mode (explicit)	Knowledge exchange mode (tacit)
Common ground between actors	Laws and predefined application forms	Newly established or updated laws	A common library of models of form models, processes and services
Communication	Editorial process	Working groups consisting of experts and service personal	Introduce model owner to the design process
Environment	Laws that change rare	Laws that change each year	Laws that change each year
Interdependency type	Federal structure	Working groups based on specific topics	External consultants

We applied the dimensions and the modes to analyze what means of coordination should be provided. First, our analysis showed a need for action in terms of introducing an editorial process as a planning mode of communication. Since each standard for an administrative procedure represents a reference model and cannot be defined in great detail, the strictly predefined process allows for keeping the standardization effective and with a minimal effort [72]. Second, our analysis shows that the feedback mode is applicable to a newly established or updated laws, since a constant update of the knowledge inside working groups is required for an effective design of a reference model Third, we came to the conclusion that a common library of models of application forms, processes and services is required as a common ground. This library would allow for reusing elements without strictly limiting the standardization process, i.e. we suggest a suitable level of standard flexibility [13]. Fourth, we figured out that a new role is needed during the design process–referred to as model designer–that should

transport existing artefact knowledge between different application domains and mediate the development of new or updated reference models.

6 Discussion and Conclusion

Our research goal was to develop a coordination theory based framework for analysis and effective standardization of administrative procedures in eGovernment. We introduced a framework that consists of three coordination modes (planning, feedback and knowledge exchange) with corresponding coordination mechanisms. Additionally, we introduce four context dimensions (common ground between actors, communication, environment, interdependency type) that help to study and describe the situation of coordination in which each mode should be applied. The framework was derived based on a single case study and interpretation of the artefacts developed in it. To exemplify the applicability of the framework, we showed how the artefacts developed in the project under are matched to particular coordination modes. Thus, we showed how the framework allows for analyzing what to coordinate towards effective standardization.

Our analysis framework represents a theoretical contribution to eGovernment research and practice [80]. Future research can comprise a quantitative study of standardization performance. Additionally, the framework can be extended with prescriptive design guidelines for an effective and efficient standardization. Practitioners can apply our framework to analyze ongoing and planned standardization activities and to evaluate the feasibility of applied coordination mechanisms.

References

1. Hanseth, O., Bygstad, B.: Flexible generification: ICT standardization strategies and service innovation in health care. Eur. J. Inf. Syst. **24**(6), 645–663 (2015)
2. Janssen, M., Charalabidis, Y., Kuk, G., Cresswell, T.: Guest editors' introduction: e-government interoperability, infrastructure and architecture: state-of-the-art and challenges. J. Theor. Appl. Electron. Commer. Res. **6**(1), I–VIII (2011)
3. Scholta, H., Balta, D., Wolf, P., Becker, J., Krcmar, H.: Standardization of service descriptions, process models and forms in public administrations: results from a survey in Germany. In: Electronic Government and Electronic Participation, pp. 245–252 (2015)
4. Becker, J., Algermissen, L., Falk, T.: Modernizing Processes in Public Administrations: Process Management in the Age of e-Government and New Public Management. Springer, Heidelberg (2012)
5. Scholl, H.J., Kubicek, H., Cimander, R., Klischewski, R.: Process integration, information sharing, and system interoperation in government: a comparative case analysis. Gov. Inf. Q. **29**(3), 313–323 (2012)
6. Radack, S.M.: More effective federal computer systems: the role of NIST and standards. Gov. Inf. Q. **7**, 37–49 (1990)
7. Borras, J.: International technical standards for e-Government. Electron. J. eGovernment **2**, 75–80 (2004)
8. Charalabidis, Y., Lampathaki, F., Askounis, D.: A comparative analysis of national interoperability frameworks. In: 15th Americas Conference on Information Systems (2009)

9. Gionis, G.A., Schroth, C., Janner, T.: Advancing interoperability for agile cross-organisational collaboration: a rule-based approach. In: Charalabidis, Y. (ed.) Interoperability in Digital Public Services and Administrations: Bridging E-Government and E-Business. Information Science Reference, Hershey (2011)
10. Balta, D., Greger, V., Wolf, P., Krcmar, H.: Why realization mismatches expectations of e-Government project benefits? Towards benefit realization planning. In: Tambouris, E., et al. (eds.) EGOV 2015. LNCS, vol. 9248, pp. 233–245. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22479-4_18
11. Balta, D., Greger, V., Wolf, P., Krcmar, H.: E-government stakeholder analysis and management based on stakeholder interactions and resource dependencies. In: 48th Hawaii International Conference on System Sciences (2015)
12. Hanseth, O., Jacucci, E., Grisot, M., Aanestad, M.: Reflexive standardization: side effects and complexity in standard making. MIS Q. **30**, 563–581 (2006)
13. Hanseth, O., Monteiro, E., Hatling, M.: Developing information infrastructure: the tension between standardization and flexibility. Sci. Technol. Hum. Values **21**(4), 407–426 (1996)
14. Bharosa, N., Lee, J., Janssen, M.: Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: propositions from field exercises. Inf. Syst. Front. **12**(1), 49–65 (2009)
15. Janssen, M.: Sociopolitical aspects of interoperability and enterprise architecture in e-government. Soc. Sci. Comput. Rev. (2011)
16. Brunsson, N., Rasche, A., Seidl, D.: The dynamics of standardization: three perspectives on standards in organization studies. Org. Stud. **33**, 613–632 (2012)
17. Lyytinen, K., King, J.L.: Standard making: a critical research frontier for information systems research. MIS Q. **30**, 405–411 (2006)
18. Markus, M.L., Steinfield, C.W., Wigand, R.T.: Industry-wide information systems standardization as collective action: the case of the U.S. residential mortgage industry. MIS Q. **30**, 439–465 (2006)
19. Rainey, H.G., Ronquillo, J.C., Avellaneda, C.N.: Decision making in public organizations. In: Handbook of Decision Making (2010)
20. Jurisch, M.C.: IT-enabled Business Process Change in Private and in Public Sector Organizations. Technische Universität München (2014)
21. Faraj, S., Xiao, Y.: Coordination in fast-response organizations. Manag. Sci. **52**(8), 1155–1169 (2006)
22. Hofmann, S., Beverungen, D., Räckers, M., Becker, J.: What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook. Gov. Inf. Q. **30**(4), 387–396 (2013)
23. Klischewski, R.: Architectures for tinkering? Contextual strategies towards interoperability in E-government. J. Theor. Appl. Electron. Commer. Res. **6**(1), 26–42 (2011)
24. Kuehn, A., Kaschewsky, M., Kappeler, A., Spichiger, A., Riedl, R.: Interoperability and information brokers in public safety: an approach toward seamless emergency communications. J. Theor. Appl. Electron. Commer. Res. **6**(1), 43–60 (2011)
25. Jacobson, D.D.: How and why network governance evolves: evidence from a public safety network. Electron. Mark. **26**(1), 43–54 (2016)
26. Gao, X., Song, Y., Zhu, X.: Integration and coordination: advancing China's fragmented e-government to holistic governance. Gov. Inf. Q. **30**(2), 173–181 (2013)
27. Spivak, S.M., Brenner, F.C.: Standardization Essentials: Principles and Practice. CRC Press, Boca Raton (2001)
28. David, P.A., Greenstein, S.: The economics of compatibility standards: an introduction to recent research. Econ. Innov. New Technol. **I**(1–2), 3–41 (1990)

29. De Vries, H.J.: Standardization: A business Approach to the Role of National Standardization Organizations. Springer, New York (1999). https://doi.org/10.1007/978-1-4757-3042-5
30. Egyedi, T.M.: Shaping Standardization: A Study of Standards Processes and Standards Policies in the Field of Telematic Services. Delft University Press, Delft (1996)
31. Cargill, C.F.: Information technology Standardization: Theory, Process, and Organizations. Digital Press, Maynard (1989)
32. Janssen, M., Snijders, B., Herkemij, F.: A Reference Architecture for Interoperable and Adaptive Processes. In: Charalabidis, Y. (ed.) Interoperability in Digital Public Services and Administrations: Bridging E-Government and E-Business, pp. 162–179. Information Science Reference, Hershey (2011)
33. Misuraca, G., Alfano, G., Viscusi, G.: Interoperability challenges for ICT-enabled governance: towards a pan-european conceptual framework. J. Theor. Appl. Electron. Commer. Res. **6**(1), 95–111 (2011)
34. Lam, W.: Barriers to e-government integration. J. Enterp. Inf. Manag. **18**(5), 511–530 (2006)
35. Hellberg, A.-S., Grönlund, Å.: Conflicts in implementing interoperability: re-operationalizing basic values. Gov. Inf. Q. **30**(2), 154–162 (2013)
36. Blum, U.: Lessons from the past: public standardization in the spotlight. Int. J. IT Stand. Stand. Res. **3**(1), 1–20 (2005)
37. Scholl, H.J., Klischewski, R.: E-government integration and interoperability: framing the research agenda. Int. J. Public Adm. **30**(8–9), 889–920 (2005)
38. Guijarro, L.: Interoperability frameworks and enterprise architectures in e-government initiatives in Europe and the United States. Gov. Inf. Q. **24**(1), 89–101 (2007)
39. Büttner, F., Bartels, U., Hamann, L., Hofrichter, O., Kuhlmann, M., Gogolla, M.: Model-driven standardization of public authority data interchange. Sci. Comput. Program. **89**, 162–175 (2014)
40. Federal Government Commissioner for Information Technology: SAGA-Modul Grundlagen Version de.bund 5.1.0. Interior FM, editor, Berlin (2011)
41. European Commission: European Interoperability Framework (EIF) for European Public Services. European Commission E, editor, Brussels (2010)
42. Galbraith, J.R.: Organization design: an information processing view. Org. Eff. Center Sch. **21**, 21–26 (1977)
43. Thompson, J.D.: Organizations in action: Social science bases of administration. Transaction Publishers, New Brunswick (1967)
44. Williams, C.K., Karahanna, E.: Causal explanation in the coordinating process: a critical realist case study of federated it governance structures. MIS Q. **37**(3), 933–964 (2013)
45. Okhuysen, G.A., Bechky, B.A.: Coordination in organizations: an integrative perspective. Acad. Manag. Ann. **3**(1), 463–502 (2009)
46. Malone, T.W., Crowston, K.: The interdisciplinary study of coordination. ACM Comput. Surv. (CSUR) **26**(1), 87–119 (1994)
47. Brown, C.V.: Horizontal mechanisms under differing IS organization contexts. MIS Q. **23**(3), 421–454 (1999)
48. Srikanth, K., Puranam, P.: Integrating distributed work: comparing task design, communication, and tacit coordination mechanisms. Strateg. Manag. J. **32**(8), 849–875 (2011)
49. Srikanth, K., Puranam, P.: The firm as a coordination system: evidence from software services offshoring. Org. Sci. **25**(4), 1253–1271 (2014)
50. Bolici, F., Howison, J., Crowston, K.: Stigmergic coordination in FLOSS development teams: Integrating explicit and implicit mechanisms. Cogn. Syst. Res. **38**, 14–22 (2016)

51. Huang, R., Zmud, R.W., Price, R.L.: Influencing the effectiveness of IT governance practices through steering committees and communication policies. Eur. J. Inf. Syst. **19**(3), 288–302 (2010)
52. Prasad, A., Heales, J., Green, P.: A capabilities-based approach to obtaining a deeper understanding of information technology governance effectiveness: evidence from IT steering committees. Int. J. Account. Inf. Syst. **11**(3), 214–232 (2010)
53. Kellogg, K.C., Orlikowski, W.J., Yates, J.: Life in the trading zone: structuring coordination across boundaries in postbureaucratic organizations. Org. Sci. **17**(1), 22–44 (2006)
54. Pawlowski, S.D., Robey, D.: Bridging user organizations: Knowledge brokering and the work of information technology professionals. MIS Q., 645–672 (2006)
55. Tanriverdi, H.: Information technology relatedness, knowledge management capability, and performance of multibusiness firms. MIS Q. **29**, 311–334 (2005)
56. Alexander, E.R.: Interorganizational coordination: theory and practice. J. Plan. Lit. **7**, 328–343 (1993)
57. March, J.G., Simon, H.A.: Organizations, pp. 34–111. Wiley, Oxford (1958)
58. Mintzberg, H.: Structure in Fives: Designing Effective Organizations. Prentice-Hall, Upper Saddle River (1993)
59. Melin, U., Axelsson, K.: Understanding organizational coordination and information systems-Mintzberg's coordination mechanisms revisited and evaluated. In: European Conference on Information Systems Proceedings (2005)
60. Myers, M.D.: Qualitative research in information systems. Manag. Inf. Syst. Q. **21**, 241–242 (1997)
61. Myers, M.D., Avison, D.E.: Qualitative Research in Information Systems: A Reader. SAGE, Newcastle upon Tyne (2002)
62. Walsham, G.: Interpreting information systems in organizations. Org. Stud. **15** (1993)
63. Klein, H.K., Myers, M.D.: A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Q. **23**(1), 67–93 (1999)
64. Goles, T., Hirschheim, R.: The paradigm is dead, the paradigm is dead…long live the paradigm: the legacy of Burrell and Morgan. Omega **28**(3), 249–268 (2000)
65. Goldkuhl, G.: Pragmatism vs interpretivism in qualitative information systems research. Eur. J. Inf. Syst. **21**(2), 135–146 (2012)
66. Cole, R., Purao, S., Rossi, M., Sein, M.K.: Being proactive: where action research meets design research. In: ICIS 2005 Proceedings, pp. 1–21 (2005)
67. Mårtensson, P., Lee, A.S.: Dialogical action research at omega corporation. MIS Q. **28**, 507–536 (2004)
68. Wilde, T., Hess, T.: Forschungsmethoden der Wirtschaftsinformatik Eine empirische Untersuchung. Wirtschaftsinformatik **49**(4), 280–287 (2007)
69. Boell, S., Cecez-Kecmanovic, D.: A hermeneutic approach for conducting literature reviews and literature searches. Commun. Assoc. Inf. Syst. **34**(1), 257–286 (2014)
70. Kotlarsky, J., Scarbrough, H., Oshri, I.: Coordinating expertise across knowledge boundaries in offshore-outsourcing projects: the role of codification. MIS Q. **38**(2), 607–627 (2014)
71. Simon, H.A.: The Sciences of the Artificial. MIT Press, Cambridge (1969)
72. Van De Ven, A.H., Delbecq, A.L., Koenig Jr., R.: Determinants of coordination modes within organizations. Am. Sociol. Rev. **41**(2), 322–338 (1976)
73. Schelling, T.C.: The Strategy of Conflict. Harvard University Press, Cambridge (1980)
74. Star, S.L., Griesemer, J.R.: Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology. Soc. Stud. Sci. **19** (3), 387–420 (1989)
75. Henningsson, S., Rukanova, B., Hrastinski, S.: Resource dependencies in socio-technical information systems design research. Commun. AIS **27**(42), 777–802 (2010)

76. Clark, H.H.: Using Language. Cambridge University Press, Cambridge (1996)
77. Reich, B.H., Benbasat, I.: Factors that influence the social dimension of alignment between business and information technology objectives. MIS Q. **24**, 81–113 (2000)
78. Kuldeep, K., van Dissel, H.G.: Sustainable collaboration: managing conflict and cooperation in interorganizational systems. MIS Q. **20**, 279–300 (1996)
79. Pfeffer, J., Salancik, G.R.: The External Control of Organizations: A Resource Dependence Perspective. Harper & Row, New York (1978)
80. Gregor, S.: The nature of theory in information systems. MIS Q. **30**, 611–642 (2006)

eLand Governance in India: Transcending Digitization

Vijeth Acharya[1], Arkalgud Ramaprasad[1,2], and Shraddha Vasudevan[1(✉)]

[1] Ramaiah Public Policy Center, Bengaluru, Karnataka, India
{vijeth.acharya, arkalgud.ramaprasad, shraddha.vasudevan}@rppc.ac.in
[2] University of Illinois at Chicago, Chicago, IL, USA

Abstract. Land governance is a complex issue to be addressed in many countries including India. All three branches of the government viz. judiciary, executive and legislature have tried to address the issue in various forms but have failed to achieve the outcomes of development. The failure, or limited success can be attributed to their narrow vision of or focus on the problem. The objectives of land governance have changed over time. With the advent of digitization, land administration has become more efficient and transparent, aiding the managerial process of land records and registration. However, the outcomes of economic, social, human, and cultural development, have been given paltry importance. In this paper we present the cases of Bhoomi and KAVERI, two initiatives to computerize land records and registration in the state of Karnataka, India. We present an ontological framework to envision the big picture in eLand governance. We argue that in achieving the desired outcome, we must see the problem in its entirety without losing focus on its key aspects. Regarding eland Governance, policy instruments, eGovernance, and land governance must work in tandem.

Keywords: eLand governance · Ontology · eGovernance
Land administration

1 Introduction

The quest to achieve better land governance in India has been a matter of continuous engagement. The aims of reforms have changed with successive governments since independence in 1947. From what started as permanent settlement of land to non-settled areas, to the protection of the rights of the landlords, to the protection of the rights of agricultural laborer and small tenants [15] and now, to the promotion of business interests for investment and job creation, the land governance regime has undergone tremendous shift. Revenue accrued from land was one of the important sources of revenue for the government in the pre-British and British era. After independence, the shift turned towards equitable distribution of land, and revenue collection did not remain as the primary objective. However, details of the land and the record maintenance pattern has not undergone much changes [32]. Land record management

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 73–84, 2018.
https://doi.org/10.1007/978-3-319-98690-6_7

has been traditionally managed by the revenue department, and with the decrease in revenue from the land the importance of keeping the updated records has gradually decreased. Added to that is the high cost of record updating with periodical surveys which is a disincentive for the department to improve the record management. Though it is an important element in bringing about the improvement in performance, accountability and transparency, record management has not been given much importance [24, 32].

With the IT (information technology) revolution in the country, many states have adopted an 'e' (digital) way to maintain the records. This shift to digital system from a traditional one has enabled better decision making. eGovernment initiatives taken by the governments are deliberated with priority owing to the optimistic sentiment of 'future in eGovernment' [1]. The transition through electronification to eGovernment can be understood by the analysis of successful and failed case studies of the implementation. eGovernment is one of the most sought after means [38] to build citizen trust in governments. eGovernment initiatives are initiated by governments to fulfil various objectives of better public service delivery, effective governance, better participation, and better management of public resources [13]. Today technology is being used for core management objectives rather than mere peripheral activity [40]. Based on the priority given to these objectives, eGovernment initiatives for land governance can either be a failure or success.

The technological impact of any eGovernment initiative can be broadly classified as optimistic, neutral, and pessimistic. The introduction of eGovernment will also result in social impacts based on the various social determinants, which is pushed by the human choices within the social structures [19].

Karnataka's Bhoomi and KAVERI (Karnataka is a state in India) initiatives were established to computerize the land records (Rights, Crops and Tenancy) of farmers and registration of the same, with all the operations that surround it [6]. Set up by the revenue department, Bhoomi was a successful eGovernment system which won several awards including example of 'best practice' by World Bank [35]. It promoted the easy retrieval of records for verification and changes which is core to the domains of computer science, information systems and public administration. Bhoomi's success is mainly attributed to (a) comprehensive background application which is inclusive of cropping details of the farmers, (b) self-sustaining nature of the project and (c) international recognition [6].

KAVERI was initiated by the Department of Stamps and Registration to computerize the registration of property transactions, lease agreements, general power of attorney, wills etc. There has been provision for integration of KAVERI with Bhoomi to solve the problem of data vacuum.

This paper critically analyses how an eGovernment initiative can promote and achieve the object of land governance. The paper analyses Bhoomi project and KAVERI of the Indian state of Karnataka through an ontological lens by analyzing, mapping, and identifying the 'bright', 'light' and 'blind' spots – the areas which have been stressed heavily, lightly, and not at all (either deliberately or accidentally).

Land governance is central to the development of nations and has become an important challenge to governance. An ideal land registry system must provide and enable secure/private/reliable/timely information/transaction/interaction for various

stakeholders in conformity with land governance principles, as encapsulated in the ontology. For better land governance, a transparent, comprehensive and legitimate record system that allows for easy mutation is also necessary. Access to land assures in investment and growth, equal opportunity, equity, and woman's empowerment [21]. Land tenure affects productivity through land related investment, using land as collateral for credit and transferring land to more productive (non-agricultural) users. Rapid economic development, expanding population, urbanization and climate change call for a strong framework to create policies regarding land use and defining clear land rights [10]. The 'e' initiatives in this area play an important role in defining the developmental outcomes as well. It is vital to note that the 'e' initiatives should act as 'means' to achieve the 'ends' of development rather than being the 'end' in themselves. This study highlights the need for envisioning the 'big picture' with the help of an ontological framework. The framework articulates large number of pathways to achieve effective land governance and achieve the outcomes with the help of eGovernance initiatives. This can be used to develop future solutions based on new and innovative pathways.

The paper is structured as follows; first, we present the ontological framework and the logic of the construction. Then we present the ontological map of Bhoomi and KAVERI. Last, we discuss the results and present the conclusion.

2 Ontology of eLand Governance

The issue of land governance and administration has been an important topic of debate in both research and policy domains. There is an oversight in understanding and articulating the architecture of the complexity of the problem [33]. The methods and efforts of solving, in most of the cases has been focused on a single part of the problem missing the 'big picture'. There is a need to bring the complexity of the problem to the foreground and address it systemically and systematically.

The ontology represents the conceptual understanding of eLand Governance [17]. With the framework, the analysis and synthesis of any domain can be carried out in a novel way [30]. The ontology can be used to deconstruct an ill-structured problem and organize the terminologies and taxonomies in a structured manner [18]. It also systematizes the description [3] of eLand Governance. The eGovernment framework [28] is further modified to encompass the land governance to form a comprehensive yet parsimonious framework. Such frameworks have been used for ontological meta-analysis and synthesis of mHealth [2] and Public Health Informatics research [31], and of research in other domains.

We present the ontological framework of eLand Governance in Fig. 1. It is a hierarchical deconstruction of the problem. The framework is constructed by an iterative process and it is scalable and extendable. (The glossary defining all the elements of the framework is not attached in the draft).

The first level of deconstruction of the problem is represented as the concatenation of Policy Instruments, eGovernance and Land Governance. It is a linguistic, logical concatenation and not an algebraic or arithmetic addition. Thus:

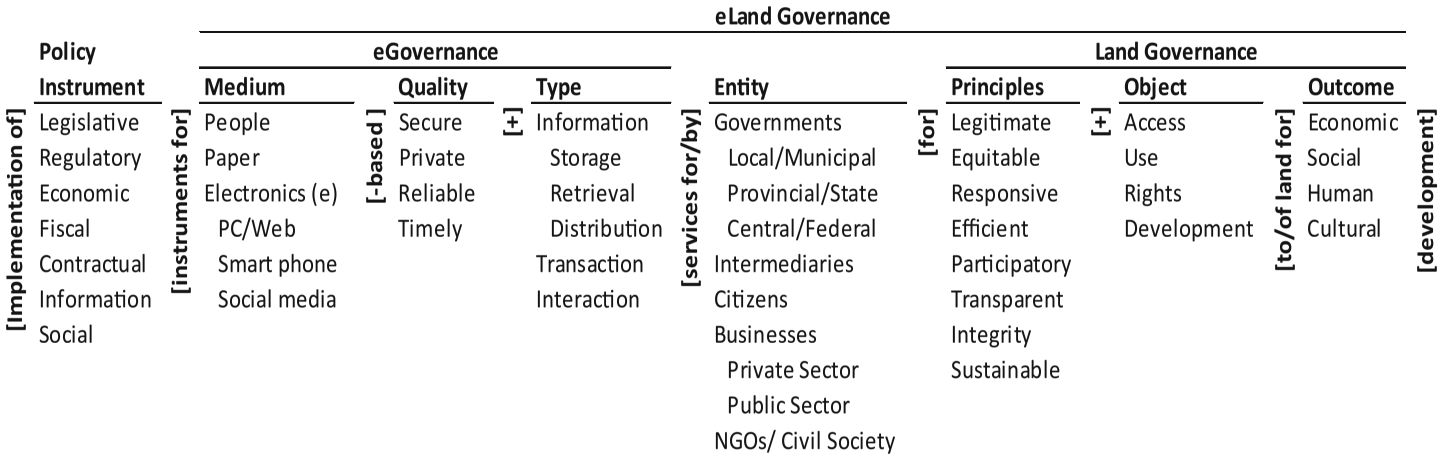

Fig. 1. Ontology of eLand governance

eLand Governance = f (Policy Instruments + eGovernance + Land Governance)

In the second level of deconstruction, eGovernance is deconstructed into Medium, Quality and Type. Land Governance is deconstructed into Principles, Object and Outcome. The dimension Entity is a common dimension for eGovernance and Land Governance. Thus:

eGovernance = f (Medium + Quality + Type + Entity)

Land Governance = f (Entity + Principles + Object + Outcome)

In the next level of hierarchy, each dimension is represented by a one level or two-level taxonomy of elements. Thus:

Policy Instruments $\subset$ (Legislative, Regulatory, Economic, Fiscal, Contractual, Information, Social)

Policy instruments are the instruments that can be utilized to have an effect on eLand Governance based on the literature on public policy [4].

Medium $\subset$ (People, Paper, PC/Web$_{(e)}$, Smart Phone$_{(e)}$, Social Media$_{(e)}$)

Quality $\subset$ (Secure, Private, Reliable, Timely)

Type $\subset$ (Storage$_{(information)}$, Retrieval$_{(information)}$, Distribution$_{(information)}$, Transaction, Interaction)

Entity $\subset$ (Local$_{(Govt)}$, State$_{(Govt)}$, Central$_{(Govt)}$, Intermediaries, Citizens, Private Sector$_{(Business)}$, Public Sector$_{(Business)}$, NGO/Civil Society)

Principles $\subset$ (Legitimate, Equitable, Responsive, Efficient, Participatory, Transparent, Integrity, Sustainable)

Object $\subset$ (Access, Use, Rights, Development)

Outcome $\subset$ (Economic, Social, Human, Cultural)

The eight dimensions arranged from left to right with connecting terms (symbols/words/phrases) forms a natural English sentence which represent potential pathways for eLand Governance. Thus, the total number of pathways in the framework are 7 * (5 * 4 * 5) * 8 * (8 * 4 * 4) = 716,800 components. Three illustrative components derived from the framework are:

Implementation of Legislative instruments for PC/Web$_{electronic}$ based timely Distribution$_{information}$ services for Citizens for Legitimate Access of land for Economic development.

Implementation of Regulatory instruments for Paper based Secure Transaction service for Private Sector$_{Business}$ for Transparent Use of land for Economic development.

Implementation of Social instruments for People based Reliable Interaction services by Local$_{Government}$ for Participatory Development of land for Social development.

The comprehensiveness of the ontology serves as the justification for its validity. It is constructed logically incorporating empirical realities. The dimensions and taxonomies are taken from the literature in eGovernance and land governance domains.

The ontology tries to capture the independence and interdependence of elements. From the analysis of interaction within the ontology, we will be able to look at the problem at different levels of granularity without forsaking the complexity of the problem, ultimately resulting in a comprehensive research. The 716,800 components encapsulated in the ontological framework represent the complete requirement to manage eLand Governance systematically and systemically. Some of the combinations may be infeasible or not instantiated. Research, policy, and practice can be mapped onto the framework to identify the frequently, infrequently and never instantiated paths for the achievement of the objective. For this study, the authors have mapped the design and objective document of Bhoomi and KAVERI projects in Karnataka [11, 12, 16, 34].

3 Results of Coding

Figure 2 below shows the results of the coding of Bhoomi and KAVERI design and objective document. The document was mapped manually onto the ontology using MS Excel using a binary scale (present, absent). Two coders reviewed the sections in the document to determine whether the element from the ontology is present or absent in the section. Each coded item was peer reviewed and final coding based on the consensus among the coders. The glossary was used to assure the validity of coding.

The number in the parentheses adjacent to each element represents the frequency of the occurrence of that element in the design and objective document. The bar below each element is a visual representation of relative frequency of the element. The minimum number is 0, if the element is not mentioned in the document. In the following, we will discuss each dimension of the ontology, left to right.

Policy Instruments: The instruments in general receive very little focus. The broad objective of both the projects is just electronification and ease of registration. However, there is a little focus on Legislative and Economic instruments.

Medium: Electronics- PC/Web is a bright spot in 'Medium' which explains that majority of the transactions happen from web, either via kiosks or from the official personnel level. There is little emphasis on paper and people in the Bhoomi and KAVERI projects.

Quality: Privacy is not at all considered in the project. Reliability and Timeliness have been given more emphasis than Security in the Implementation.

Type: All the elements in Service type attracts emphasis in the implementation of Bhoomi and KAVERI. The highest concentration is on Transaction followed by Information Storage and Interaction.

Entity: Local Government and Citizens are highly emphasized in the project. Businesses and Civil Societies attract very little attention. Though Intermediaries play an important role in smooth functioning of the systems, their emphasis in design and objective has been light.

Principles: Legitimate, Transparent and Efficient Principles are dominant, whereas Responsive, Equitable and Integrity are very lightly focused. Participatory and Sustainability principles of land governance are not at all looked into in the projects.

Object: Access and Rights are dominantly focused neglecting the Use and Development of land.

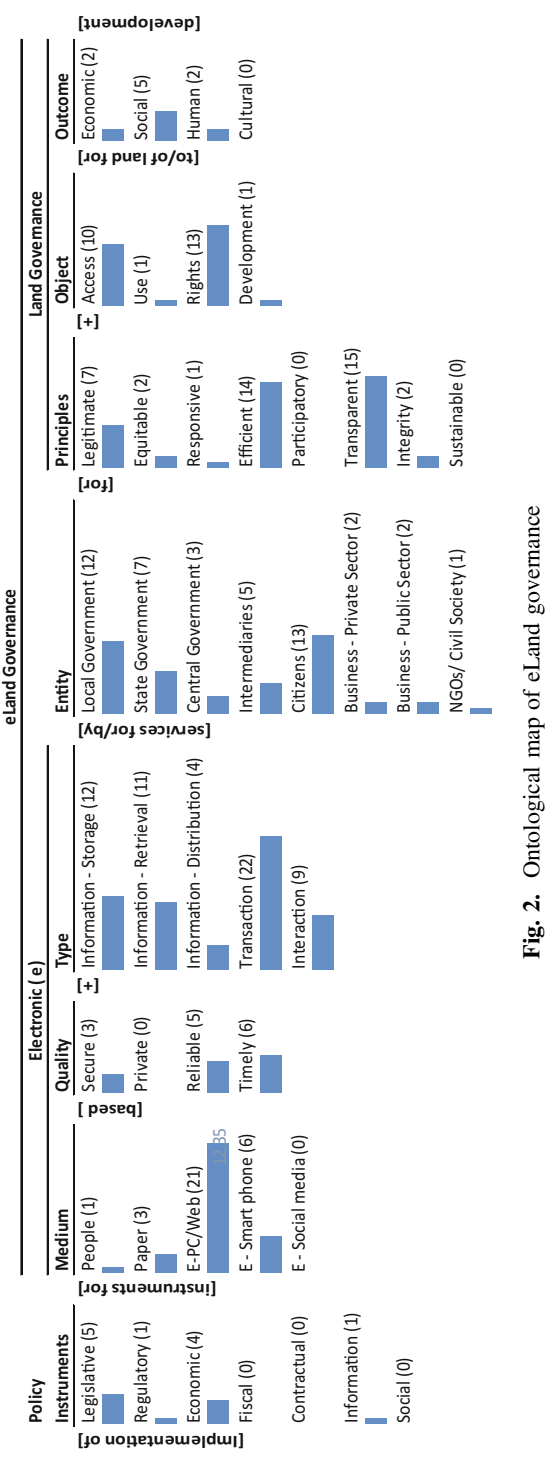

Fig. 2. Ontological map of eLand governance

Outcome: The Outcomes of Land Governance has not been the primary focus of the project. Economic, Social and Human development outcomes are very lightly emphasized while Cultural Development aspect has been an oversight.

4 Discussion

Land information needs to be carefully managed and handled to obtain the potential benefits. The program envisioned in the Sixth Five Year Plan (From 1980–1985) for the records updating found a strong base in the Seventh Five Year Plan which said "Land records form the base for all land reform measures and therefore, regular periodic updating of land records is essential in all states." [36]. Bhoomi and KAVERI have succeeded in the process of automation of land records and registration. However, the extent to which these eGovernment initiatives have assured in comprehensive land governance remains to be seen. Our mapping reveals that Bhoomi and KAVERI, as information systems, have only mildly aided in comprehensive land governance to bring about economic, social, human, and cultural development. The implicit idea of achieving development outcomes by computerization (ibid) forms a flawed logic in a diverse country like India which has a complex land administration which varies considerably across states [7], and without proper planning and evidence based policy actions, the achievement of outcomes is difficult.

While digitization may be the first step towards eLand Governance, it alone cannot ensure developmental outcomes. An efficient land administrative system requires an effective land policy framework along with an interoperable (semantic, legal, inter – community and technical) land information infrastructure [22, 39]. The Bhoomi and KAVERI systems, established in the 2000s, haven't undergone any changes to further their objectives beyond mere automation. While they have served as successful land information systems in some capacity, they have thus far acted as standalone systems that have failed to integrate into a larger land governance framework to realize broader, more tangible outcomes.

Land administrative systems must strive to achieve social political and environmental sustainability through government decision making and citizen participation [14]. However, the blind spot under the sustainable development outcome of the framework, reveals that the KAVERI and Bhoomi, as information systems have not informed the overall land administrative system to attain sustainable development. Although Bhoomi is a sustainable information system, generating revenue to finance itself, it has not facilitated sustainable land governance principle.

The mapping shows that KAVERI and Bhoomi provide legitimate documents that ensure land rights and access. Clear rights and access leads to security of land tenure which promotes greater incentive to invest in land and use it for productive activities [8]. However, this land security has not led to any substantial economic development outcome. Lack of proper field verification has led to incorrect entries into the Bhoomi system [32] as well. Therefore, there exists a presumed security of rights by land registration systems only effective on paper, without any practical value [9].

The Bhoomi and KAVERI system have assisted the governments from the local to the central level in decreasing order, as shown by our mapping results. While the

managerial process has been taken care of, how this land information system has assisted the local Karnataka government in facilitating development outcomes through principled land governance, remains an unanswered question. This could be a result of narrow objectives, initially set for the systems. If the objectives of these systems are not prioritized efficiently, the result of economic, social, human and cultural development may never be realized.

Eminent domain and other restrictions on land use, exercised by governments may lead to tenure insecurity and discourage investment unless such powers are applied in a transparent manner [9]. In practice, distribution of landownership, doesn't happen through market mechanisms but instead through an already existing power structure which is further solidified by the land registration system (ibid). Furthermore, outdated land records have been transferred on to the Bhoomi system without proper field verification [32]. Therefore, appropriate legislative and regulatory instruments for enforcing property rights must work in tandem with error free, verified information systems to ensure development outcomes and prevent misuse of land.

The results of our coding reveal a blind spot on Participatory Land governance. The principle of participation in land governance has been overlooked by Bhoomi and KAVERI. To obtain timely land information results, a participatory land information system is required [25]. Such a system requires the consideration of various social, political and economic aspects, apart from technical solutions and administrative procedures (ibid). Furthermore, a participatory approach to land use planning accounts for the rights of all vulnerable communities [10] thereby promoting human development.

The People element under the medium dimension has also been only mildly emphasized. KAVERI aims at 100% automation of the registration process while Bhoomi aims at disintermediation, and complete automation of land record management. The presence of People in the framework is essential as no digitized system can operate without some degree of the human element [1]. Even the most advanced forms of eGovernment use the element of People [29] as a medium for delivering services. In the context of developing countries, it becomes necessary to fit in human intermediaries to ensure appropriate penetration of ICT technology to reach the poor [27] as well as to endorse participatory land governance. Using Volunteered Geographic Information has helped in securing land rights through a participatory manner and has in turn facilitated social and human development of the rural people in the villages of Africa [25].

Although Bhoomi was initiated for the easy retrieval of land records by farmers at nominal rates, farmers demands were unaccounted for, in the designing the system. The ability of technology to change the attitudes of the citizens is derived from their trust in the government [38]. In essence, Bhoomi, has not facilitated active participation of citizens in social change, but has merely created an efficient information management model [37].

The digitization of land information has led to transparency and ease of transactions which have helped greatly in administration [26]. This in turn has helped in transparent land governance. However, through the process of automation, Bhoomi initiative has essentially aided the revenue department, in its bureaucratic functions. For comprehensive e governance initiatives, focus on demand side as to why and when citizen use eGovernance facilities becomes as important as supply side perspectives [20].

5 Conclusion

Land information systems are an indispensable aspect of land governance systems. Bhoomi and KAVERI have successfully computerized land records and registration and facilitated the easy storage, retrieval and updating of data, while citizen benefits and development outcomes have taken up a secondary priority. While technology remains a driving force for land governance, the objectives of such governance need to been renewed in order to attain overall developmental outcomes [5]. The latest development in managing land registries is the use of blockchain technology. This technology provides for a decentralized, public ledger that is seen as a solution to the challenges of record keeping [23]. Bhoomi and KAVERI initiatives haven't been updated since their commencement. A new approach must be put forward to prioritize developmental outcomes, rather than assuming such outcomes will be realized merely through digitization. A modern land governance system must function to guarantee land rights, tenure and credit, provide information for land planning and development, and manage land transfer and assets, [14] which can enable societal development. This requires synchrony between a strong policy framework, capable land administration and a robust information system. Therefore, information systems must not work as silos (as in the case of KAVERI and Bhoomi) but instead work in harmony with the policy and administration system. Bhoomi and KAVERI need to function within the policy framework of land governance and in turn assist in providing informed evidence-based policies of land governance. The ontology reveals an absence of emphasis on the participatory and sustainable principle, a very mild emphasis on equitable, responsive and integrity principles, and a substantial emphasis on the efficiency principle of land governance. This is because the multidimensional approach of effecting comprehensive land governance through information systems is absent. Land use planning and development must also take place through a participatory approach through the inclusion of people in the information systems to achieve sustainability. A narrow objective of digitization without focusing on the bigger picture can lead to a myopic vision thereby under-utilizing the capacity of Bhoomi and KAVERI. ICT must be applied as an enabler for principled land governance to achieve the outcomes of economic, social, cultural and human development. Using the ontology of Land eGovernance, pathways for attainment of comprehensive development through land governance aided by ICT infrastructure can be uncovered.

References

1. Acharya, V., et al.: eBhoomi - an ontological analysis. In: National Seminar on Emerging trends in Public Policy and implications for development'. Jyoti Nivas College, Bengaluru (2018)
2. Cameron, J.D., et al.: An ontology of and roadmap for mHealth research. Int. J. Med. Inform. **100**, 16–25 (2017)
3. Cimino, J.J.: In defense of the desiderata. J. Biomed. Inform. **39**(3), 299–306 (2006)
4. Crespi, F., et al.: Taxonomy of implemented policy instruments to foster the production of green technologies and improve environmental and economic performance (2015)

5. Dale, P.: Is technology a blessing or curse in land administration (1999)
6. De', R.: E-government systems in developing countries: stakeholders and conflict. In: Wimmer, M.A., Traunmüller, R., Grönlund, Å., Andersen, K.V. (eds.) EGOV 2005. LNCS, vol. 3591, pp. 26–37. Springer, Heidelberg (2005). https://doi.org/10.1007/11545156_3
7. Deininger, K.: A Strategy for Improving Land Administration in India, no. 33. World Bank (2008)
8. Deininger, K., et al.: Identifying and addressing land governance constraints to support intensification and land market operation: evidence from 10 African countries. Food Policy **48**, 76–87 (2014)
9. Deininger, K., Feder, G.: Land registration, governance, and development: Evidence and implications for policy. World Bank Res. Obs. **24**(2), 233–266 (2009)
10. Deininger, K.W., et al.: The Land Governance Assessment Framework: Identifying and Monitoring Good Practice in the Land Sector. World Bank, Washington, D.C. (2012)
11. Department of Land Resource Ministry of Rural Development, GOI: Success Stories on National Land Records Modernization Programme. Deity, GOI (ND) (2014)
12. Satya Murty, D., et al.: KAVERI- Karnataka Valuation and E-registration (2002)
13. Elkadi, H.: Success and failure factors for e-government projects: a case from Egypt. Egypt. Inform. J. **14**(2), 165–173 (2013)
14. Enemark, S., et al.: Building modern land administration systems in developed economies. J. Spat. Sci. **50**(2), 51–68 (2005)
15. Gae, R.S.: Land law in India: with special reference to the constitution. Int. Comp. Law Q. **22**(2), 312–328 (1973)
16. Government of Karnataka: Note on Integration of KAVERI with Banks. Department of Stamps and Registrations (2014)
17. Gruber, T.R.: Ontology. In: Liu, L., Özsu, M.T. (eds.) Encyclopedia of Database Systems. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-39940-9_1318
18. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing? Int. J. Hum.-Comput. Stud. **43**(5–6), 907–928 (1995)
19. Heeks, R., Bailur, S.: Analyzing e-government research: perspectives, philosophies, theories, methods, and practice. Gov. Inf. Q. **24**(2), 243–265 (2007)
20. Helbig, N., et al.: Understanding the complexity of electronic government: implications from the digital divide literature. Gov. Inf. Q. **26**(1), 89–97 (2009)
21. International Bank for Reconstruction and Development: India Land Governance Assessment Nation Synthesis Report (2015)
22. Kalantari, M., Rajabifard, A.: An interoperability toolkit for eLand Administration (2005)
23. Lemieux, V.L.: Trusting records: Is Blockchain technology the answer? Rec. Manag. J. **26** (2), 110–139 (2016). https://doi.org/10.1108/RMJ-12-2015-0042
24. Mokhtar, U.A., Mohammad Yusof, Z.: Electronic records management in the Malaysian public sector: the existence of policy. Rec. Manag. J. **19**(3), 231–244 (2009)
25. Moreri, K.K., et al.: Volunteered geographic information quality assessment using trust and reputation modelling in land administration systems in developing countries. Int. J. Geogr. Inf. Sci. **32**, 1–29 (2018)
26. National Resource Cell for Decentralized District Planning: E-Governance for Information Dissemination Bhoomi, Karnataka (n.d.)
27. Rajalekshmi, K.G.: E-governance services through telecenters: the role of human intermediary and issues of trust. Inf. Technol. Int. Dev. **4**(1), 19–35 (2007)
28. Ramaprasad, A., Sánchez-Ortiz, A., Syn, T.: An ontology of eGovernment. In: Tambouris, E., et al. (eds.) EGOV 2015. LNCS, vol. 9248, pp. 258–269. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22479-4_20

29. Ramaprasad, A., et al.: Gaps in local eGovernment research: an ontological analysis. In: Proceedings of the 21st Americas Conference on Information Systems (AMCIS 2015), Puerto Rico, USA (2015)
30. Ramaprasad, A., Syn, T.: Design thinking and evaluation using an ontology. In: Helfert, M., Donnellan, B., Kenneally, J. (eds.) EDSS 2013. CCIS, vol. 447, pp. 63–74. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13936-4_6
31. Ramaprasad, A., Syn, T.: Ontological meta-analysis and synthesis. Commun. Assoc. Inf. Syst. **37**(7), 138–153 (2015)
32. Saxena, N.C.: Updating land records: is computerisation sufficient? Econ. Polit. Wkly. **40**, 313–321 (2005)
33. Simon, H.A.: The architecture of complexity. Proc. Am. Philos. Soc. **106**(6), 467–482 (1962)
34. Singh, P.: Best Practices Study on Land Resources Management in India. Centre for Rural Studies, Lal Bahadur National Academy of Administration, Mussoorie, India (2016)
35. Solomon, B., et al.: Bhoomi: 'E-Governance', Or, An Anti-politics Machine Necessary to Globalize Bangalore? (2007). https://casumm.files.wordpress.com/2008/09/bhoomi-e-governance.pdf
36. Thakur, V., et al.: Land management system in India past, present and future (2010). https://www.geospatialworld.net/article/land-management-system-in-india-past-present-and-future/
37. Thomas, P.: Bhoomi, Gyan Ganga, e-governance and the right to information: ICTs and development in India. Telemat. Inform. **26**(1), 20–31 (2009)
38. West, D.M.: E-government and the transformation of service delivery and citizen attitudes. Public Adm. Rev. **64**(1), 15–27 (2004)
39. Williamson, I., Enemark, S.: Land Administration for Sustainable Development (2009)
40. Yildiz, M.: E-government research: reviewing the literature, limitations, and ways forward. Gov. Inf. Q. **24**(3), 646–665 (2007)

Coordinating Public E-services - Investigating Mechanisms and Practices in a Government Agency

Fredrik Söderström(✉), Ulf Melin, Ida Lindgren, and Zara Galzie

Linköping University, 581 83 Linköping, Sweden
{fredrik.soderstrom, ulf.melin, ida.lindgren, zara.galzie}@liu.se

Abstract. Coordination is a critical enabler when creating and managing coherent, integrated, secure and smart public electronic services (e-services) as a part of digitalization. With an increased demand for such services, coordination as an internal organizational phenomenon is becoming increasingly important. Based on a qualitative case study, and informed by coordination theory, this paper investigates two different theoretical views applied on internal e-service coordination within a government agency in Sweden. At the outset, the agency is seeking one generic way to coordinate the current heterogeneous and fragmented internal e-service landscape in a more efficient way. Hence, our aim also includes investigating the prerequisites and potential for this type of coordination. We conduct this study in two stages. First, we apply a well-established theoretical lens from organizational theory on a set of coordination efforts, thereby perceiving coordination as a planned and anticipated activity based on a fixed set of mechanisms. Second, we apply a lens of coordinating as emergent practice, which allows for an in-depth investigation of more flexible and dynamic aspects of coordinating activities in daily work. By combining these two views, we argue that this approach can facilitate and increase understanding of the dynamics and flexibility needed to understand the type of coordination needed in public e-service contexts. This can also imply that there is no single best practice or 'one-size-fits-all' approach to internal e-service coordination. Instead, organizations need to acknowledge the need for multi-dimensional views revealing the inherent complexity of coordination; as planned as well as emerging activities.

Keywords: Coordination · Coordination mechanisms · E-government
E-services · Digital services · Public sector ICT · Public sector digitalization

1 Introduction

This paper elaborates on the formal mechanisms and emerging practices of coordination of electronic services (e-services) in a government agency. Coordinating activities are recognized as vital parts of organizing and refer to the actions taken by humans in organizations or other social settings to generate anticipated and appropriate outcomes:

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 85–97, 2018.
https://doi.org/10.1007/978-3-319-98690-6_8

"To organize is to assemble ongoing interdependent actions into sensible sequences that generate sensible outcomes." [1, p. 3]. Coordination is an essential subset of organizing with fundamental characteristics of formalizing actions and reducing undesired variation by increased control and the ability to anticipate actions [2–4]. Coordination also minimizes the risk of overlapping and even conflicting actions among stakeholders [5]. Organizational relationships enacted via informal, dynamic activities need to be aligned [6] in turn creating a tension between anticipated and spontaneous activities of coordination; described as the organizational paradox [4, 7]. Achieving a balance between reductions of variation and the need for flexibility therefore becomes critical. Transdisciplinary coordination studies describe organizational actions as mutually dependent hence, balanced handling is required [4, 8]. Acknowledging this need for balance, we use different streams of coordination theory as our basis for formulating two different theoretical lenses; viewing coordination as formal mechanisms [e.g. 3] and coordinating as emerging practices [e.g. 6].

In a public e-service context, we agree with researchers putting forth the need for fixed as well as emergent views of coordination, i.e. formal mechanisms and informal practices [e.g. 6]. As artefacts, e-services are intertwined with policies and practices of coordinating such services [8, 9]. This creates a need to investigate and understand the characteristics, complexity and coordination of public e-services, in policy, as formal mechanisms, as well as in emergent practice. Benefits and expectations such as increased internal efficiency and external availability [e.g. 10, 11] make the ability to provide e-services a necessity for a majority of public organizations. Due to Sweden's governance model, citizens communicate with several government agencies on a regular basis. However, due to public organizations' independence, the level of digitalization, e.g. regarding e-services, varies considerably across the public sector. As exemplified by the case study, complexity and fragmentation are significant challenges to be handled in a digitalization context, and these challenges can also be related to coordination efforts [12].

Investigating e-service coordination challenges is thereby still of importance to government digitalization practice, as well as e-government research. This paper highlights different shapes and roles of coordination in an e-service context by investigating its formal and informal dimensions and expressions. This paper addresses the following research questions (RQ): (1) How are e-services coordinated in a public agency? (2) How can this coordination be described using two different theoretical lenses of coordination? (3) What can we learn from combining these lenses to develop a more dynamic and flexible view of e-service coordination? To address these questions, we must first investigate how e-services are coordinated at the agency (RQ1), then review how the two different theoretical lenses can be applied (RQ2) and finally, how these lenses can be combined and applied for interesting insights (RQ3). Therefore, the aim of this paper is to present an approach in which formal as well as informal acts of coordination are taken into consideration. This paper is organized as follows; the second section covers related research such as coordination as mechanisms, coordinating as practice and e-service coordination. In the third section, the research approach and case study are presented. The analysis and findings are outlined in the fourth section followed by a discussion in section five. The paper is concluded with some concluding remarks and suggestions for future research.

2 Related Research

2.1 Coordination as Mechanisms

As introduced above, well-known views on coordination and related mechanisms are presented by scholars such as March and Simon [2] and Mintzberg [3]; Mintzberg [13]. Three activities are described as central: coordination through *standardization*, coordination through *planning*, and coordination through *feedback* [2]. Similarly, Van de Ven et al. [14] describe three predominant modes of coordination as *impersonal* by plans or programming, *personal* by mutual adjustments or feedback and *group-based* by unscheduled or scheduled activities. Partly based on March and Simon [2], Mintzberg develops the following set of coordination mechanisms: (1) *mutual adjustment*, (2) *direct supervision*, (3) *standardization of skills and norms*, (4) *work processes*, and (5) *results* [3, 13].

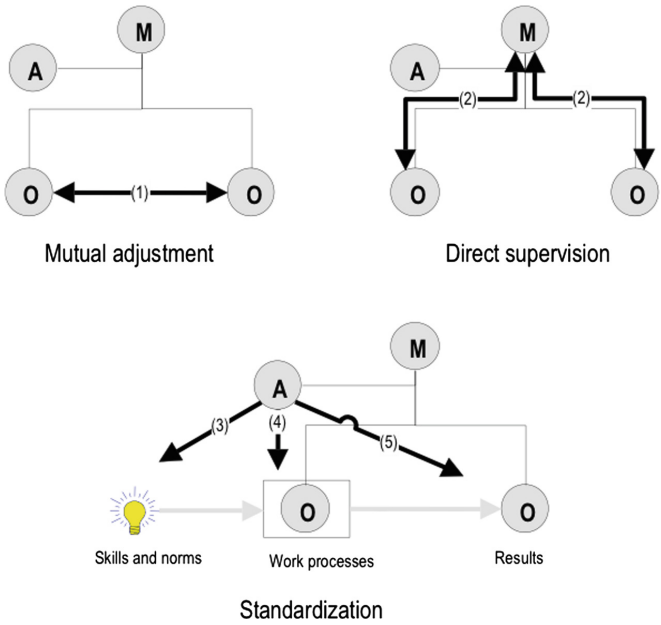

Fig. 1. Coordination mechanisms *[3 revised by Melin and Axelsson 2005]*

Mutual adjustment (1) achieves coordination of activities by the process of informal coordination. The control of the activities rests mostly in "the hands of the doers" on an operative level ("O" in Fig. 1). **Direct supervision** (2) expresses a more hierarchical model where coordination is achieved by assigning one actor to a role responsible for the work of others ("M" in Fig. 1). In this role, managers typically issue instructions to actors on the operative level thus monitoring these actions becomes important. An analyst role ("A" in Fig. 1) can coordinate organizational activities with

different types of **standardization** (3, 4, and 5) indicating a specification or programming of the contents of work processes (4). Outputs can be standardized by specifying the **results** of the work (5), for example, the dimensions of an e-service or the output of coordinating it. **Skills** (3) are standardized when the expertise required to perform work is specified beforehand [3]. **Norms** (3) are standardized to influence human action; as an indirect, or even subtle, form of coordination [13] (cf. organizational culture). In addition to the described hierarchical or vertical coordination in the form of direct supervision and standardization, horizontal or non-hierarchical coordination is often performed by a coordinator role reporting to top-management but with no formal authority in the areas of coordination [15].

We acknowledge some of the criticisms concerning coordination as expressed above, put forward by, e.g. Larsson [16] and Melin and Axelsson [17] implying coordination being too focused on planned activities, designed elements of organizing and material flow in organizations. Further, Adler [18] describes the importance of the temporal dimension of coordination since its mechanisms and organizational dependencies need to change over time. In response, to achieve a dynamic and flexible view of coordination, we combine the lens of coordination as mechanisms with the view of coordinating as practice described below.

2.2 Coordination as Practice

Okhuysen and Bechky [19] describe the interdisciplinary field of coordination research as increasingly interested in the processes and practices of coordination. This enables more in-depth investigations of coordinating as an ongoing dynamic aspect of organizational activities [20]. Faraj and Xiao [21] emphasize the areas of expertise and dialogic coordination hence aspects such as the community of practice and knowledge sharing become critical to the distributed expertise needed for coordinating. Thus, coordinating as practice refers to enactment in accordance with formal mechanisms as well as an activity emerging in the absence of such mechanisms. This implies that there is not one optimal way of coordination since coordinating activities are focused on managing dependencies in organizational settings [8]. Dynamic coordination often takes place during problem-solving tasks among organizational participants [22]. In this paper, we apply the conceptual model of Jarzabkowski et al. [6] consisting of the five overlapping stages or cycles of coordinating as practice (Fig. 2). **Enacting disruption** (1) focuses on any obstacles or disruptions of barriers forced upon coordinating activities caused by formal organizational policies and structures and **orienting to absence** (2) represents the attempts to organize or re-organize any coordinating activities needed by employees to be able to perform their operational work. **Creating elements** (3) starts when actors initiate the formation activities to facilitate the needed coordination. Elements that are further developed and refined during **forming patterns** (4). **Stabilizing patterns** (5) occurs when elements and patterns stabilize as an acknowledged practice.

Accepting and acknowledging these elements and patterns of coordination as appropriate ways of performing tasks occurs during the final stage of this cycle. Unlike the described formal mechanisms, this five-cycle model emphasizes aspects such as disruptions of work processes and the absence of coordination enabling us to seek

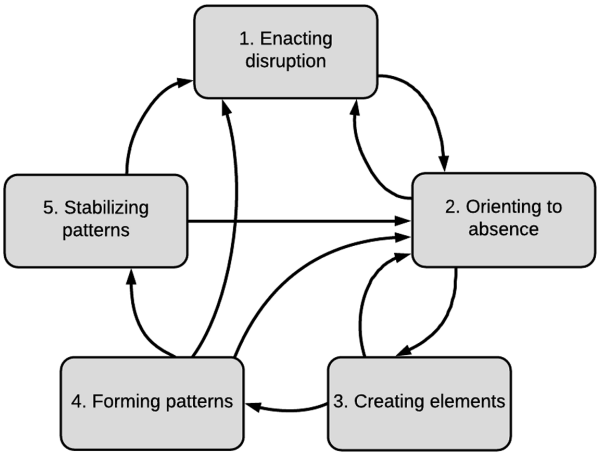

Fig. 2. Cycles of coordinating as practice [adapted from 6, p. 919]

relationships between the formal and informal dimensions of coordination. In previous research, we identify a lack of application of the view of coordination as practice; hence, this is the theoretical gap we address in this paper.

2.3 Public E-services and Coordination

Using public e-services as a communication channel between citizens and government is perceived and promoted as a way of improving the efficiency and effectiveness of public sector organizations [11]. Public e-services are electronically mediated services, provided by public organizations, through which the users and the supplying organization co-create some value through the users' consumption of the service [23]. Policies, as well as citizens' needs and expectations, govern these services. There are persisting challenges associated with public e-service development such as the provision and use; service quality [24], service design [25], and uptake and use [26]. Significant challenges stem from the inherent complexity of this phenomenon. E-services refer to a process as well as an IT-artefact [23] which in turn challenges how we can understand and organize related work. Moreover, public e-services vary in complexity and type [27, 28] with different types of services putting different demands on the organization providing them. We thereby stress the need to take this complexity as well as coordination of e-services into account. Furthermore, coordination and control facilitate and increase the efficiency of delivery of services in public bureaucratic organizations [29]. Public sector digitalization also increases the need for coordination due to higher levels of technical complexity, service complexity as well as organizational inter- and intra-dependencies. Coordination of services is therefore required to maintain high levels of quality due to the increasing number of dependencies related to areas such as services, channels, and systems [12]. There are studies in e-government inspired by formal coordination (above) for example that focused on inter-organizational e-service integration [e.g. 30]. However, the research related to

e-service coordination seems to be limited with exceptions such as coordination of integrated service delivery across public organizations [30] and dependencies in multi-channel service delivery [12].

3 Research Approach and Case Study

The presented research is based on a qualitative and interpretive case study [31, 32] conducted at the Swedish Transport Administration (STA) as part of a research project investigating the development and use of e-services within this agency. Conducted during 2016–2017, the overall project aimed to gain a deeper understanding of how public organizations can organize e-service development and implementation ensuring value for internal as well as external stakeholders. The STA is responsible for the long-term planning of the national transport system covering all types of traffic, as well as building, operating, and maintaining public roads and railways. The agency is located in several geographic regions in Sweden and divided into departments each responsible for development and provision of its own e-services. At the outset of this study, the STA called for one generic way to coordinate their heterogeneous and fragmented e-service landscape. We interpret this fragmentation as the result of historically weak coordination and control in this area, in combination with high levels of independence among departments within the agency.

Our qualitative data collection included two different sources. First, nine semi-structured interviews [33] were conducted from October to December 2017 with respondents from different departments. Two interviews were conducted face-to-face, with the rest conducted via telephone. Open-ended questions guided the interviews and focused on different shapes and roles of coordination. Since this study aims to investigate existing as well as emergent or even lack of coordination of public e-services, this approach enabled all performed interviews to contribute positively to the study. For reasons of space, only a selection of empirical findings are presented and analysed in this paper. Second, a hermeneutic literature review [34] was performed to increase our understanding of coordination in the form of mechanisms and emerging practices as well as coordination in a public e-service context. Therefore, during this study, the process of identifying and interpreting relevant literature was a continuous process affected by our developing knowledge of the focused research field. This study is thereby based on empirical data while using theory as a guide [32]. Data analysis was conducted iteratively during data generation, and responses were categorized, as a part of a content analysis approach [35]. Emerging categories were identified in previous research as an example of a reflexive research process [36]. To be able to model mechanisms and practices of coordination based on respondents' statements, the organizational and operational contexts were analysed in detail. This enabled the identification of formal as well as informal structures and activities of coordination and the lack thereof.

4 Coordinating E-services at the STA

Below we present the analysis of e-services within the STA, based on the two selected theoretical lenses of coordination. The aim of the first stage of analysis (Sect. 4.1) is twofold. First, we want to categorize the area of e-service coordination through the lens of coordination as mechanisms. Second, this stage serves as a stepping-stone providing us with crucial findings to perform the next stage of analysis. The aim of the second stage of the analysis (Sect. 4.2) is to investigate the coordinating practice indicated by areas of mutual adjustment identified during the previous stage. For the second stage, we use an excerpt of empirical data collected at the Customer Service department.

4.1 E-service Coordination Mechanisms

The **E-service Catalogue** is located at the external website and provides a single point of access to the majority of the e-services provided by the agency. Services range from simple web-forms to advanced forms of e-services. Without any explicit direct supervision, the Communications department was responsible for bringing together all e-services as owned and provided by different departments into a central e-service catalogue. Hence, this created a need for interdepartmental coordination across the organization. As an example of *mutual adjustment*, the Communications department investigated and assessed the status of all e-services as well as located responsible departments and contact persons. Another example is the **Customer Service** department with caseworkers assigned to the tasks of handling incoming client cases via channels such as telephone and email. We perceive caseworkers as forced to take the responsibility of forming coordination elements and patterns of their own to be able to solve the cases at hand. Public agencies, as well as other government bodies, are requested to have a precisely formulated strategy on how to transpose the benefits of modern technology into government business, i.e. public sector digitalization. We, therefore, interpret the agency's **Digitalization strategy** as an example of *direct supervision* as driven by central strategies and policies on national as well as organizational level. However, we identify a gap between policy and operational level due to challenges of translating the aims of the policy into actual benefits for local operations.

The mechanism of *standardization* is identified in current efforts regarding service design, the **Service Forum**, aimed at applying a standardized and aligned perspective and organization to assess and develop different kinds of services across the organization. The **Service Forum** is also identified as an example of *direct supervision* since we interpret this forum as an arena for increased coordination and operationalization of a service design model across the organization. However, with this forum potentially affecting processes as well as results across the organization, we have not seen any convergence between this coordination mechanism and structures and processes related to the internal handling of e-services within the agency. The **Case Management Project** is another example of an effort with distinct aims of providing coherent technical infrastructure for case handling across the organization hence, this project relates to technical *standardization*. Still, in its early stages, the project will require extensive and explicit coordination during its phases of design and implementation to provide the anticipated high level of standardized and automated coordination of cases.

Moreover, this project will most likely have a significant impact on the internal handling of cases in turn linked to external e-services towards citizens.

4.2 E-service Coordinating Practices

We begin by expanding the scope of **enacting disruption** to include organizational contexts not covered by any formal mechanisms of coordination. Our interpretation of disruption marks a need for informal coordinating practices to solve the challenges at hand. The following statement describing the e-service landscape at the STA identifies this type of disruption. *"From our perspective, it's very messy and complex. We wish to have a complete picture; that it should work in the same way everywhere."* (Customer Service, interview). Caseworkers at the Customer Service department are struggling to handle the challenges of supporting a large number of different e-services, and the external e-service catalogue has had a rather unanticipated impact on this department regarding how client cases are forwarded internally. *"Today it probably ends up at Customer Service. It's our contact number showing on the website and e-service forms, but it's just a service catalogue, really. When we receive the case, we handle it to be best of our ability."* (Customer Service, interview). The e-service catalogue serves as a single point of access and therefore, from the client's view, creates an expectation of Customer Service being a similar single point of contact. **Orienting to absence** frequently occurs since caseworkers regularly face challenges needing the expertise of other employees across the organization. Caseworkers start orienting by trying to locate appropriate contact points across the organization; interpreted as an investigative and time-consuming effort.

A result of being an informal act of coordination, the lack of formal support, routines and procedures also results in new challenges. *"Sometimes we detect problems in e-services, so we contact the Communications department, but they do not feel they want to act on the matter and the IT department feels they do not own this case since this is a form that goes to someone else. Therefore, we end up being caught in the middle. We identify problems that arise, but nobody wants to take responsibility."* (Customer Service, interview). Instead of finding the appropriate contact point to report and solve these problems, different departments are engaged in negotiations about what party is responsible for a particular e-service. One assumption is that it would be in the Customer Service's self-interest to try to develop a higher level and more formal way of coordinating practice shared among caseworkers. However, findings point out the challenges of creating these more structured elements. *"So the biggest challenge is to keep track of what e-services are available and which one you should refer to and what to do with the different types of questions we receive"* (Customer Service, interview). Thereby, at Customer Service, we identify the cycle of **creating elements** as caseworkers establish the needed forms of coordination but so far, we have not been able to find any indications of these practices reaching beyond this particular cycle of creating elements. The cycles of **forming and stabilizing patterns** do not seem to apply in this empirical context. Instead, existing patterns seem to prescribe that the coordinating practice needs to be formed and re-formed on a case-by-case basis. *"We have no list at the office showing contact persons for each e-service. We simply make contact with*

someone within the organization and investigate further. A lot of our work is done in this manner." (Customer Service, interview).

5 Discussion

Our study shows the presence of the two suggested forms of coordination from previous research within the STA; as *planned mechanisms* and as *emerging practices* in daily work. The agency applies different types of *direct supervision* and *standardization* in their efforts to formalize and anticipate actions [2–4]. We motivate this by how e-services are presented on the external website (E-Service Catalogue) and how cases are internally managed with the support from the new technical infrastructure (Case Management Project) potentially increasing internal efficiency as well as the external availability [e.g. 10, 11]. During the first stage of analysis, we identify *mutual adjustment* existing on the operative level [3] in contexts such as the E-service Catalogue and the caseworkers' handling of client cases in the Customer Service department. This coordination takes place in areas with very limited, or total absence of, direct supervision and standardization with a clear aim of problem-solving [22]. Coordinating activities thereby include characterizing dependencies and identifying potential coordination activities [8] as well as locating the needed expertise [21] across the organization. In contrast to the previously described *analyst role* in formal mechanisms acting upon coordination in the form of standardization, we perceive caseworkers as operative level *coordination analysts* or *informal coordinators* supporting activities based on mutual agreements [c.f. 3]. These activities are efforts to mitigate the effects of the absence or lack of formal coordination mechanisms.

To develop a better understanding of these mutual adjustments as areas of emerging coordinating practices, we use an empirical subset of data as the basis of stage two of the analysis. During the first cycle of *enacting disruption*, caseworkers describe the fragmented and heterogeneous e-service landscape as having a significant impact on their daily work preventing them from having a clear picture of the internal e-services linked to the external e-service catalogue. To solve client cases, caseworkers need the distributed knowledge from different departments across the organization such as Communications and IT, as well as specific departments in their roles as e-service providers. However, there is a lack of an internal overview of structures and dependencies such as internal case handling systems linked to external e-services. This indicates a lack of understanding of the different roles of e-services; as artefacts, as well as interlinked processes [23]. We also interpret this as a result of the complexity and fragmentation in the organization [12] combined with a lack of knowledge and support regarding organizational interdependencies behind these e-services. Thereby, caseworkers need to bridge this gap on a daily and operational basis, a gap to which we, in this context, assign the concept of *disruption*.

During the cycle of *orienting to absences*, caseworkers start to assess and investigate further the situation in their efforts to deconstruct the problem at hand into manageable elements. Since cases regarding e-services do not follow any organizational structures or boundaries [12], caseworkers try to identify dependencies to be able to perform activities of horizontal coordination in each case. Since there is no clear

picture of the e-service landscape, they usually approach an actor deemed suitable as the first contact for the continued investigation. When clients report problems of a more general nature on the website, caseworkers approach internal support functions such as the Communications and IT departments. It is therefore interesting to learn that these support functions seem to avoid taking any responsibility and act upon such cases. This can be interpreted as another indication of a lack of knowledge and understanding of the e-service landscape as well as a consequence of the caseworkers' lack of authority in their roles as coordinators [15].

When *creating elements*, caseworkers depend on their individual experiences and tacit knowledge of dependencies linking the external e-service to internal department functions to facilitate the creation of the needed coordination in each case. This is also an example of the agency's lack of acknowledgment and support of caseworkers at Customer Service in their roles as *informal coordinators*. Acting without any support of formal coordination, such as direct supervision [3], this prevents the development of any *stabilized and formalized patterns* to support further coordinating activities. Without any formal acknowledgment and support, coordinating practices take place, but in a significantly un-coordinated way. From our point of view, caseworkers at Customer Service have a vital role as *informal coordinators*; a role that needs be both formally acknowledged and supported by the organization.

The main findings of this study are the following three: First, by addressing RQ1, we confirm that activities related to internal structures and processes related to e-services are coordinated in several ways ranging from direct supervision and standardization to mutual adjustment and emerging coordinating practices. Findings show several informal coordination activities performed without the informal coordinator, e.g. caseworkers at Customer Service department, being aware of underlying interdependencies between different structures and processes. This results in a disability to support other coordinating activities, as well preventing any forming and stabilizing patterns of recurring coordinating activities [6]. Second, by answering RQ2, we describe and analyse the identified contexts and occurrences of coordination where the first analytical stage focuses on *coordination mechanisms* thus guiding the second stage where a deeper understanding of *coordinating as practice* is developed. Hence, this approach enables us to uncover and investigate the complexity and interdependencies of coordinating mechanisms as well as coordinating practice in daily activities. Third, as a result of this study, we put forth the need to expand existing formal oriented views of coordination, such as fixed sets of coordination mechanisms, with perspectives that can potentially identify and capture emerging informal dimensions of coordinating. By addressing RQ3, we thereby agree with scholars such as Malone and Crowston [8] and Jarzabkowski et al. [6] emphasizing a balanced handling of organizational actions and a need to consider and understand coordination as mechanisms as well as practice.

6 Concluding Remarks and Future Research

In this study, we apply two views on coordination to empirical data to further investigate and analyse formal and informal dimensions of coordination in a fragmented and heterogeneous public agency e-service context. Although the agency was seeking *one*

generic way or a 'one-size-fits-all' approach, e.g. a framework or best practice, to coordinate internal e-services, our findings point towards the need of also considering more *balanced, dynamic and flexible views of coordination* in this area. By applying an emergent view on coordination, we also agree with Okhuysen and Bechky [19] stating that "coordination is an ongoing accomplishment in organizations." (p. 493). In this paper we, therefore, conclude that the e-services to be coordinated do not depend on technical dimensions of the artefacts, formal structures and mechanisms alone; they are also intertwined with organizational and social constructs inherent to the accepted working practices and informal relationships and dependencies among participating actors [c.f. 8, 9]. The implications of this study assumes a need to look beyond pre-formulated fixed-set view and instead increase our sensitivity to the emerging dynamics of coordination in public sector digitalization. As our final conclusion in this paper, we argue that being too focused on a single or formal views of coordination will delimit the organization from a broader in-depth understanding of the informal dimensions of coordinating as emerging activities.

We put forth the added value of combining approaches of coordination that cover planned as well as emerging forms of coordination, in research as well as in practice. This combination of views is in agreement with Jarzabkowski et al. [6] emphasizing the need for looking beyond fixed and ready to be used views of coordination. We are fully aware that various governance approaches, such as IT-governance frameworks, can address challenges of internal coordination of public e-services. However, such frameworks focus on formal management in terms coordination and direction of IT-related decisions. As a contrast, we stress the importance of understanding the informal dimension of coordinating as practice on the organizational level. We argue that this will improve the understanding of the objects of coordination (e-services), as well as its organizational context. However, emergent coordinating activities can be challenging to identify while placing formal coordination in the foreground. There is also an inherent risk of bias towards the anticipated outcome of coordination as known a priori.

Regarding the limitations of this study, it could be argued that it was performed during a limited timeframe with a rather small sample size in one single organization hence, the level of generalizability should be considered low. However, as stated in the introduction, given our study's qualitative and interpretive character, our aim is not to provide any statistically generalizable results, but rather put forward the need of a better understanding of different shapes and roles of coordination by presenting and applying one potential approach where formal mechanisms as well as informal practices of coordination can be potentially integrated, an provide an illustrative case with analytical generalizability. Since this is our first attempt bringing these two theoretical lenses of coordination mechanisms and coordinating practices together, we seek to further develop our understanding of the complexity and interwoven character of technical, organizational and social dimensions of internal coordination, in an e-government context in general and in an e-service context in particular. A particular area of interest in future research is to further develop the suggested approach of combining and mutually adjusting or coordinating, formal coordination and informal coordinating activities for an increased understanding of the different shapes and forms of coordination.

Acknowledgements. This work was supported by the Swedish Transport Administration.

References

1. Weick, K.: The Social Psychology of Organizing, 2nd edn. McGraw-Hill, New York City (1979)
2. March, J.G., Simon, H.A.: Organizations. Wiley, New York (1958)
3. Mintzberg, H.: Structure in Fives: Designing Effective Organizations. Prentice Hall, New Jersey (1983)
4. Thompson, J.D.: Organizations in Action: Social Science Bases of Administrative Theory. Transaction Publishers, Piscataway (1967)
5. Thomas, G., Botha, R.A., van Greunen, D.: Understanding the problem of coordination in a large scale distributed environment from a service lens view—towards the South African public sector e-Administration criteria for coordination support. Gov. Inf. Q. **32**(4), 526–538 (2015)
6. Jarzabkowski, P.A., Lê, J.K., Feldman, M.S.: Toward a theory of coordinating: creating coordinating mechanisms in practice. Org. Sci. **23**(4), 907–927 (2012)
7. Melin, U.: The enterprise system as a part of an organization's administrative paradox. J. Enterp. Inf. Manag. **23**(2), 181–200 (2010)
8. Malone, T.W., Crowston, K.: The interdisciplinary study of coordination. ACM Comput. Surv. (CSUR) **26**(1), 87–119 (1994)
9. Orlikowski, W.J., Scott, S.V.: 10 sociomateriality: challenging the separation of technology, work and organization. Acad. Manag. Ann. **2**(1), 433–474 (2008)
10. Axelsson, K., Melin, U., Lindgren, I.: Public e-services for agency efficiency and citizen benefit—findings from a stakeholder centered analysis. Gov. Inf. Q. **30**(1), 10–22 (2013)
11. Bertot, J., Estevez, E., Janowski, T.: Universal and contextualized public services: digital public service innovation framework. Gov. Inf. Q. **33**(2), 211–222 (2016)
12. Klievink, B., Janssen, M.: Coordinating e-government service delivery. In: 11th Annual International Conference on Digital Government Research dg.o 2010, pp. 209–216. Digital Government Society of North America, Puebla (2010)
13. Mintzberg, H.: Five Ps for strategy. In: Mintzberg, H., Quinn, J.B., Ghoshal, S. (eds.) The Strategy Process. Prentice-Hall, Upper Saddle River (1998)
14. Van de Ven, A.H., Delbecq, A.L., Koenig Jr., R.: Determinants of coordination modes within organizations. Am. Sociol. Rev. **41**, 322–338 (1976)
15. Lawrence, P.R., Lorsch, J.W.: Differentiation and integration in complex organizations. Adm. Sci. Q. **12**(1), 1–47 (1967)
16. Larsson, R.: Coordination of action in mergers and acquisitions – interpretive and systems approaches towards synergy. Ph.D. thesis. Lund University, Sweden (1990)
17. Melin, U., Axelsson, K.: Understanding organizational coordination and information systems-Mintzberg's coordination mechanisms revisited and evaluated. In: Proceedings of the 13th European Conference on Information Systems, pp. 1634–1644 (2005)
18. Adler, P.S.: Interdepartmental interdependence and coordination: the case of the design/manufacturing interface. Org. Sci. **6**(2), 147–167 (1995)
19. Okhuysen, G., Bechky, B.: Coordination in organizations: an integrative perspective. Acad. Manag. Ann. **3**(1), 463–502 (2009)
20. Feldman, M.S., Orlikowski, W.J.: Theorizing practice and practicing theory. Org. Sci. **22**(5), 1240–1253 (2011)

21. Faraj, S., Xiao, Y.: Coordination in fast-response organizations. Manag. Sci. **52**(8), 1155–1169 (2006)
22. Ching, C., Holsapple, C.W., Whinston, A.B.: Reputation, learning, and coordination in distributed decision-making contexts. Org. Sci. **3**(2), 275–297 (1992)
23. Lindgren, I., Jansson, G.: Electronic services in the public sector: a conceptual framework. Gov. Inf. Q. **30**(2), 163–172 (2013)
24. Jacobsen, D.I.: Adopting and refining e-services—the role of organization size. Public Org. Rev., 1–13 (2016)
25. Goldkuhl, G.: What does it mean to serve the citizen in e-services?-towards a practical theory founded in socio-instrumental pragmatism. Int. J. Public Inf. Syst. **3**(3), 135–159 (2007)
26. Angelopoulos, S., Kitsios, F., Kofakis, P., Papadopoulos, T.: Emerging barriers in e-government implementation. In: Wimmer, M.A., Chappelet, J.-L., Janssen, M., Scholl, H. J. (eds.) EGOV 2010. LNCS, vol. 6228, pp. 216–225. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14799-9_19
27. Jansen, A., Ølnes, S.: The nature of public e-services and their quality dimensions. Gov. Inf. Q. **33**(4), 647–657 (2016)
28. Lindgren, I., Melin, U.: Time to refuel the conceptual discussion on public e-services – revisiting how e-services are manifested in practice. In: Janssen, M., et al. (eds.) EGOV 2017. LNCS, vol. 10428, pp. 92–101. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64677-0_8
29. Cordella, A., Tempini, N.: E-government and organizational change: reappraising the role of ICT and bureaucracy in public service delivery. Gov. Inf. Q. **32**(3), 279–286 (2015)
30. van Os, G.: The challenge of coordination: coordinating integrated electronic service delivery in Denmark and the Netherlands. Inf. Polity **16**(1), 51–61 (2011)
31. Myers, M.D.: Qualitative Research in Business & Management. SAGE Publications, London (2009)
32. Walsham, G.: Interpretive case studies in IS research: nature and method. Eur. J. Inf. Syst. **4** (2), 74–81 (1995)
33. Myers, M.D., Newman, M.: The qualitative interview in IS research: examining the craft. Inf. Org. **17**(1), 2–26 (2007)
34. Boell, S.K., Cecez-Kecmanovic, D.: A hermeneutic approach for conducting literature reviews and literature searches. Commun. Assoc. Inf. Syst. **34**(1), 257–286 (2014)
35. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology, 2nd edn. SAGE Publications, Thousand Oaks (2004)
36. Alvesson, M., Sköldberg, K.: Reflexive methodology: new vistas for qualitative research. SAGE, London (2009)

The War on Corruption: The Role of Electronic Government

João Martins[1,2(✉)], Bruno Fernandes[2], Ibrahim Rohman[1], and Linda Veiga[2]

[1] UNU-EGOV, Rua de Vila Flor 166, 4810-445 Guimarães, Portugal
joao.martins@unu.edu
[2] University of Minho, Braga, Portugal

Abstract. This paper explores the relationship between electronic government and corruption. Our main hypothesis is that digital government can increase the availability of information on public sector, making it more difficult to engage in corrupt activities without being caught. Our estimations are performed under several alternative methodologies: ordinary least squares, first differences, fixed effects and random effects. Based on a dataset that covers more than 150 counties, we find that a higher E-Government Development Index is associated with better corruption outcomes. Our results are robust to all the methodologies that we have implemented. The quality of telecommunication infrastructures and human capital can amplify the impact of digital government in corruption. We investigate how this relationship varies across income levels. Upper middle income countries are the ones that can extract more benefits from developing digital government. Low income countries should invest in infrastructures and education in order to benefit from electronic government.

Keywords: Corruption · E-Gov · Information · Online services Transparency

1 Introduction

Corruption is largely perceived as an obstacle to a country's development and good functioning of the institutions [27]. Through the last two decades, Information and Communication Technologies (ICT) in general, and Electronic Government (e-gov) in particular, have been pointed as efficient and convenient for reducing corruption [6]. E-gov is perceived as capable of increasing transparency and accountability, as well as of promoting several Sustainable Development Goals (SDGs) [13].

The inclusion of e-gov and ICT-based solutions in governments' programmes is worldwide a trend [32]. Digital government's potential to reduce corruption was highlighted in the 2003 Global Corruption Report: it "reduces the discretion and opportunity for arbitrary action available to civil servants when dealing with applicants on a case-by-case basis. Moreover, as the possibility of exposure of wrongdoing is enhanced, the fear of consequent embarrassment can be a deterrent to corrupt practices." ([7]: 24); "It increases chances of exposure by maintaining detailed data on transactions, making it possible to track and link the corrupt with their wrongful acts." ([7]: 30).

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 98–109, 2018.
https://doi.org/10.1007/978-3-319-98690-6_9

There is already an extensive literature on the influence of e-gov on socioeconomic outcomes. A meta-analysis of the literature on digital government's impacts [3] shows that it helps to control corruption. In this literature, empirical studies using econometric techniques, are scarce. They suggest that e-gov has an effect on the size of the shadow economy [34], on the ease of doing business [4] and on corruption control [2].

To the best of our knowledge, there are only two econometric studies [2, 26] that investigate the relationship between e-gov and corruption. Additionally, these studies used data which does not cover the last ten years. Therefore, more studies are necessary and, given the fast speed of innovations in e-gov, it is urgent to analyse more recent data. Using a panel that goes from 2005 to 2016, and covers more than 150 countries, we assess digital government's impact on corruption. Our study improves on the previous literature by using the indexes that are produced by the United Nations (UN) as a measure of e-gov. These are currently the benchmark on measuring e-gov development, being quite comprehensive in terms of content and coverage.[1] We also extend the literature by exploring how this relationship differs across income groups, which is particularly helpful when extrapolating policy implications from the results.

The remaining of the paper is structured as follows. Section 2 reviews the literature about corruption, focusing on its economic relevance and main determinants, and highlights the potential role of e-gov on fighting corruption. Section 3 describes the data. Section 4 explains the empirical methodology. Section 5 presents and discusses the empirical results. Finally, Sect. 6 concludes.

2 Literature Review

Corruption is considered one of the most pervasive global problems. Countries with high levels of corruption are usually poor, and the majority of them don't have sustainable economic growth [21]. This global problem is also associated with political instability, coups d'état, politicians assassinations and less freedom of speech [9].

Government's structure and organization influences corruption and bribery in different manners and is often used to justify differences in the corruption levels among countries. [27] claimed that competition reduces bribery when government goods (e.g. passports, import permissions) are provided by agencies, since theft can be diminished. To maintain the revenues from bribery, secrecy is important for the elite, because it prevents the entry of newcomers, who bring changes and innovation. The lack of competition also seems a common explanation for the development of bribery behaviour [1]. The openness to foreign investment and antitrust regulation also prevent corruption, since less competition leads to higher rents for existing firms, and higher bribes for bureaucrats who have control rights over those firms.

The degree and tenure of democracy are also deeply connected with government's structure and organization. Mature democracies tend to exhibit lower levels of corruption while the reverse happens in new democracies due to fragile institutions and free entry into the collection of bribes by the agencies that provide public goods [27].

[1] See [25] for a survey of e-gov indexes.

As a democratic pillar, the freedom of press is seen as a good mechanism to lower corruption, because journalists are less prone to hide corruptive actions [9].

Political connections also tend to be stronger in more corrupt countries. On a panel of countries, [14] concluded that political connectedness is associated with more corruption, less press freedom, and a worse legal environment. In Pakistan, [18] found that public banks favour politically connected firms in terms of loans. They lend more money to firms that have at least a politician in their board, even if those firms have higher rates of default. For Uganda, [30] found that profitable firms and firms with low bargaining power have to bribe more to obtain licenses, public services or tax reductions.

Computers have been used in governmental agencies almost since their beginning, but the emergence of the term e-gov and of e-gov as an academic field were phenomena of the late 1990's [15]. Since then, several e-gov conferences and specialized journals were created by the academic community, but a definition of electronic government that is accepted as standard by the literature is yet to be formulated [13]. An example of an early definition of e-gov, is that "e-Government refers to the process of connecting citizens digitally to their government in order that they might access information and services offered by government agencies" ([19]: 89).

A more holistic concept of digital government was proposed by [16], who classified the e-gov concept as an evolution-like concept, evolving towards more complexity and contextualization. According to [16], four stages in the concept evolution process can be identified: digitization or technology in government, transformation or electronic government, engagement or electronic governance and contextualization or policy-driven electronic governance. According to [33], the ICT-based solutions in government that can contribute to the reduction of bureaucracy, informal economy and corruption go until the fourth stage of the evolution process.

The relationship between e-gov and corruption can be established through a political economy approach [22]. Since electronic records are easier to store and access, e-gov can facilitate audits, preventive checks and ongoing investigation of corrupt acts. Moreover, e-gov can contribute towards greater interoperability and integration of public services, which is likely to increase the probability of detecting a corrupt official. Additionally, the open government and the open data initiatives can promote transparency and collaboration [20], contributing to the control of corruption.

Several case studies argue that e-gov initiatives can be used to reduce corruption. A well-documented case is the OPEN system in the Seoul metropolitan area.[2] Implementing surveys to more than 11000 citizens and 2000 city officials, [10] concluded that 49% of the citizens and 45% of the officials thought the initiative was helpful to reduce corruption. Survey methodologies were also used to conduct case studies in Ethiopia, Fiji and Bangladesh: [8, 23, 24] concluded that e-gov can diminish corruption and improve the relationship between citizens and government.

[2] "OPEN" is an acronym for the "Online Procedures Enhancement," which consisted in a web-based service to transact civil applications for permits, registrations, procurements, contracts, and approvals, among other matters of service to citizens by the Seoul Metropolitan Government and its district offices.

As mentioned before, empirical studies about the impact of e-gov on corruption using either cross country or panel samples are scarce, particularly those that use an independent variable that attempts to measure e-gov.[3] [26] regressed the West's e-gov efficiency index on the Corruption Perceptions Index reported by Transparency International. They concluded that e-gov is positively correlated with better corruption outcomes. Using a panel of 149 countries, with 2 observations per country, [2] combined a first-differences approach with instrumental variables. It argues that e-gov, measured by the Brown University index, positively affects the control of corruption, especially in non-OECD countries. Besides e-gov, media freedom and GDP *per capita* also turned out to positively influence the control of corruption.

3 Data

Our proxy for the development of e-gov in a given country is the UN E-Gov Development Index (*EGDI*). This index is released biannually and is based both on the E-Gov Surveys implemented by UNDESA and on secondary data from the International Telecommunications Union (ITU) and UNESCO. The most recent edition was in 2016, with a coverage of 193 countries [32]. The *EGDI* is calculated as an arithmetic average of three sub-indexes: the Online Services Index (*OSI*), the Telecommunications Infrastructure Index (*TII*), and the Human Capital Index (*HCI*).

The *OSI* is the sub-index that is more intrinsically related with e-gov, no matter the e-gov definition that we use, and therefore, receives particular attention in our analysis. This variable is constructed based on a questionnaire, which seeks to evaluate several e-gov related features of each country's national portal, as well as other governmental portals, and websites of the ministries of education, labour, social services, health, finance and environment.[4] The *TII* is based on data provided by ITU and measures the development of telecommunication infrastructures. The *HCI* is based on educational related indicators provided by the UNESCO.

The dependent variable of our regressions is the Control of Corruption Index (*CCI*). The *CCI* is a variable from the Worldwide Governance Indicators (WGI), which proxies perceptions of the extent to which the public power is exercised for private gain, (taking into account both grand and petty forms of corruption), and of the degree to which the state is captured by elites and private interests.[5]

Besides the e-gov related indexes, other explanatory variables are included in our regression models. Regarding economic variables, we use the log of the GDP *per capita* (*loggdp*) and the degree of openness of the economy (*openness*). GDP is

[3] [11] concluded that more access to information can contribute to reduce corruption. However, the authors used the digital access index, as an independent variable.

[4] The evaluated features include online service delivery, whole of government approaches, open government data, multi-channel service delivery, e-participation mobile services, usage up-take, digital divide and innovative partnerships through the use of ICT. Both the availability of the e-tools and the easiness of the interaction are taken into account. See [32] for a detailed description of the methodology.

[5] For a detailed description of the methodology see [17].

considered by several studies, e.g. [31], to be the strongest predictor of corruption. The degree of openness of the economy is measured by the sum of exports and imports over GDP. As more open economies are more exposed to external competition, we expect them to be more transparent and less corrupt [1]. We also take into account institutional factors. Following [9], we include in our set of independent variables the index of press freedom (*pressfree*) by the Reporters without Borders. Contrary to what happens for the other indexes, a higher value of *pressfree* signals a worse outcome. Following [12], we admit that bureaucracy can influence corruption. Our proxy for the level of bureaucracy is the index of the Ease of Starting a Business (*startbus*), from the World Bank's Doing Business Project. In this index, a higher score means that it is easier to start a business. Finally, we also take into account the World Bank's income classification of the country. We do this because the e-gov's impact on corruption may differ across income groups. High income countries have already a high score in the *CCI* and their margin for additional progress is small. Low income countries may not have a critical mass of infrastructures and human capital that allows them to benefit from e-gov. Table 1 illustrates the differences in the mean values of the variables included in the analysis, according to the four World Bank's income classifications: low income (LI), lower middle income (LMI), upper middle income (UMI) and high income (HI). As can be seen from the table, higher income levels are associated with higher levels of e-gov, less corruption, more freedom of the press, less bureaucracy, and more open economies.

A graphic inspection of the data allows us to visualize the relationship between e-gov and corruption. Figure 1 shows the correlation between *EGDI* (horizontal axis) and *CCI* (vertical axis), in 2016. There is clear evidence of a positive correlation between the two variables. As expected, developed countries tend to appear in the upper part of the graph, where *EGDI* and *CCI* are both higher.

4 Methodology

Our empirical methodology can be divided into three main parts. In the first, we analyse the relationship between e-gov and corruption in the year of 2016. In the second one, we implement a panel data analysis. At last, we evaluate how the relationship between e-gov development and corruption varies across income groups.

We start by estimating the model presented in Eq. (1) by OLS:

$$CCI_{i,2016} = \beta_0 + \beta_1 . EGov_{i,2016} + \gamma . X'_{i,2015} + e_{i,2016} \quad (1)$$

where CCI stands for the CCI of a given country i in 2016, and EGov represents the value of our main independent variables, either the EGDI or the OSI, in country i. X' represents a vector of lagged control variables, γ the vector of coefficients associated with the control variables and e the error term. Vector X' includes the following variables: openess, loggdp, pressfree, and startbus.

Although our baseline OLS models are aligned with the corruption literature, we cannot exclude the possibility biasness in the estimators. Panel data techniques allow us to control for time and country level time-invariant effects, which may be a source of

Table 1. Main Variables' average by income level (2016). (Originally the *EGDI* and the *OSI* are defined in a scale from 0 to 1, whereas the *CCI* is in a scale from −2.5 to 2.5. To facilitate the interpretation of the regression coefficients and to have all indexes in a similar scale, we rescaled them from 0 to 100).

	LI	LMI	UMI	HI	Overall
EGDI	21.7	38.2	50.6	73.0	49.3
OSI	20.6	36.9	45.1	70.0	46.4
CCI	31.9	38.9	44.0	71.2	49.9
Openness	65.2	76.4	83.7	117.2	89.4
Pressfree	34.8	41.4	37.8	23.3	34.8
Startbus	71.3	80.6	81.8	89.2	82.0

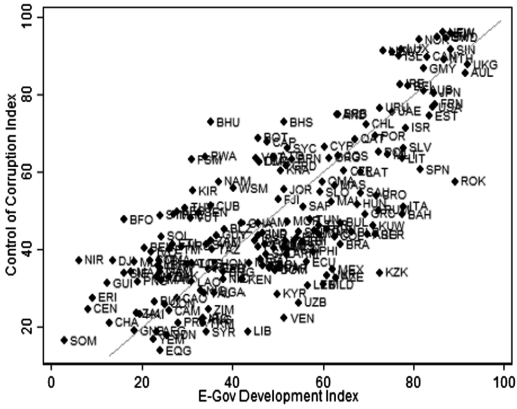

Fig. 1. Correlation between *CCI* and *EGDI* in 2016. Country codes and a linear fit are included.

such biasness. Our panel covers the years of 2005, 2008, 2010, 2012, 2014 and 2016. We start by testing if our results are robust to the First Differences (FD) approach. This method explores the variation in the variables within the time horizon of the panel [35]. In the FD regressions we compute the difference between the final and the initial value of each variable for each country. The final value is the value of 2016, while the initial value is the value of 2005. Equation (2) represents our FD approach.

$$\Delta CCI_i = \alpha + \beta_1 . \Delta EGov_i + \gamma . \Delta X_i' + \Delta e_i \tag{2}$$

where $\Delta CCI_i = CCI_{i,2016} - CCI_{i,2005}$, $\Delta EGov_i = EGov_{i,2016} - EGov_{i,2005}$, $\Delta X_i' = X_{i,2016}' - X_{i,2005}'$ and $\Delta e_i = e_{i,2016} - e_{i,2005}$.

We move forward by estimating Fixed Effects (FE) and Random Effects (RE) regressions, which can be represented by Eq. (3):

$$CCI_{i,t} = \beta_0 + \beta_1 EGov_{i,t} + \gamma . X_{i,2015}' + \lambda_t + \mu_i + \varepsilon_t \tag{3}$$

where, t represents the year. λ_t are time effects, which in our specification are captured by year dummy variables, and μ_i are country fixed or random effects depending on

which model we use. After estimating our models using both FE and RE, we perform Hausman tests to evaluate which model is the most appropriate.[6] In the FE and RE regressions we use a smaller set of control variables to avoid multicollinearity problems. High Variance Inflated Factors (VIF) were obtained in the loggdp and in the easestartbus variables.[7] As multicolinearity can damage the statistical power of the analysis and the stability of the significance of the coefficients of the model, these variables were removed from the regressions. Therefore, in this case, our X' vector is composed only by the lagged variables openness and pressfree.

Finally, we estimated separate regressions for the four World Bank's income classifications. Countries' different contexts must be taken into account when implementing e-gov strategies [29]. As the *EGDI* and the *OSI* indexes evaluate the same features for the entire world, we may expect their impact on corruption to differ across groups of countries with different income levels.

5 Results

5.1 Cross Section Results

We start by describing the estimation results obtained through the OLS method. Table 2 reports these results. In column (1) the EGDI is the main explanatory variable, while in column (2) the OSI is used.

As can be seen from Table 2, the e-gov proxy variables, as well as the *loggdp* and the *pressfree,* are always statistically significant at the 1% significance level. The magnitude of the estimated impact of e-gov is higher when we use the *EGDI* than when the *OSI*, because *EGDI* includes not only *OSI* but also *TII* and *HCI*. This can be interpreted as follows: e-gov services are important to mitigate corruption, but their impact can be amplified by the quality of human capital and telecommunication infrastructures. The OLS models predict that a one point increase in the *EGDI* is associated with an increase in the score of the corruption index of 0.27 points. The same increase in the *OSI* is associated with an increase of 0.18 in the *CCI*. The estimated coefficient for *loggdp* is higher when we use the *OSI* than when we use the *EGDI* as explanatory variable. This can be explained by the fact that when we use the *OSI* the levels of the telecommunication infrastructures and human capital are removed from the model. Finally, *Openess* did not turn out as statistically significant and *easestartbus* was marginally significant in model (2).

5.2 Panel Results

Panel data regressions allow to control for country and time-invariant factors. We test if our results are robust to FD and FE or RE estimations. Hausman tests revealed that, FE

[6] When the test results in a negative chi-square value, the FE model is preferable.

[7] As we estimated regressions for different income groups, we must take into account the constant term. It give us the base levels of the CCI in different regions. Therefore, following [5], we also look at the uncentered VIF values when testing for multicollinearity.

Table 2. OLS results. *Notes:* All models were estimated with a constant; Robust standard errors in parentheses; Statistical significance: *** p < 0.01, ** p < 0.05, * p < 0.1

Variables	(1)	(2)
	CCI	CCI
EGDI	0.269**	
	(0.103)	
OSI		0.182***
		(0.053)
l.openess	0.023	0.030
	(0.019)	(0.018)
l.loggdp	4.683***	5.560***
	(1.388)	(0.951)
l.startbus	0.110	0.110*
	(0.067)	(0.064)
l.pressfree	−0.416***	−0.437***
	(0.067)	(0.065)
Observations	154	154
Adj_r2	0.727	0.735

is preferable to RE. The results are reported in Table 3. The FD results are in columns (1) and (2) and the FE results in columns (3) and (4).

In general, the results presented in Table 3 are similar to those reported in Table 2. However, the magnitude of the estimated coefficients are now lower than those obtained by OLS. The coefficients associated with the *EGDI*, the *OSI* and the *pressfree* variables have the same sign and are still statistically significant, but their magnitude is smaller. Once again the coefficients associated with the *EGDI* are higher than the coefficients associated with the *OSI*. The FD regression predicts that a one point increase in the *EGDI* from 2005 to 2016 was associated with a 0.1 point increase in the *CCI*, while the same increase in the *OSI*, during the same period, was associated with a 0.07 point increase in the *CCI*. The FE estimator suggests that a one point increase in the *EGDI* is associated with an improvement of 0.071 in the *CCI*, while the same increase in the *OSI* is associated with a 0.04 points increase in the same variable.

5.3 Results by Income Group

Our final empirical exercise consists on evaluating the differences in the relationship between e-gov and corruption across different income groups. Table 4 reports the results obtained for the four levels of the World Bank's income classification.[8]

Table 4 shows that the *EGDI* variable is statistically significant in the LI and UMI groups of countries, while the *OSI* variable is only statistically significant in the UMI

[8] For parsimonious reasons, in this table we do not report the standard deviations of the coefficients neither the coefficients of the lagged control variables *openness* and *pressfree*. These results will be provided by the authors upon request.

Table 3. First differences and fixed effects results. Notes: All models were estimated with a constant; Robust standard errors in parentheses; Statistical significance: *** p < 0.01, ** p < 0.05, * p < 0.1

Variables	(1)	(2)	Variables	(3)	(4)
	ΔCCI	ΔCCI		CCI	CCI
ΔEGDI	0.104* (0.059)		EGDI	0.071** (0.031)	
ΔOSI		0.072** (0.030)	OSI		0.037** (0.016)
Δopeness	0.029 (0.020)	0.031 (0.020)	l.openess	−0.001 (0.011)	−0.002 (0.011)
Δloggdp	10.955*** (2.415)	10.520*** (2.401)			
Δstartbus	−0.010 (0.033)	−0.011 (0.032)			
Δpressfree	−0.072* (0.039)	−0.070* (0.039)	l.pressfree	−0.044*** (0.016)	−0.044*** (0.016)
Method	FD	FD		FE	FE
# of countries	127	127		168	168
Observations	127	127		950	950
adj_r2	0.189	0.208	Overall_r2	0.644	0.549

Table 4. Fixed and Random Effects results by income groups of countries. *Notes:* All models were estimated with time dummies for the years of our panel: 2008, 2010, 2012, 2014 and 2016 (2005 is the base category); Statistical significance: *** p < 0.01, ** p < 0.05, * p < 0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CCI	CCI	CCI	CCI	CCI	CCI	CCI	CCI
Group	LI	LMI	UMI	HI	LI	LMI	UMI	HI
EGDI	0.26**	0.08	0.12***	0.02				
OSI					0.10	0.04	0.07***	0.01
β_0	30.0***	31.8***	39.7***	71.0***	33.4***	33.6***	41.9***	71.4***
Model	RE	RE	RE	FE	FE	RE	RE	FE
#Country	29	45	47	47	29	45	47	47
Obs	159	257	262	272	159	257	262	272
r2	0.157	0.173	0.228	0.059	0.225	0.18	0.209	0.054

group of countries. This finding is consistent with the idea that it is necessary to achieve a critical level of telecommunications infrastructures and human capital before starting to benefits from e-gov. It also suggests that the significance of the *EGDI* coefficient in the LI countries relies largely on the components of this index other than the *OSI*. In turn, the UMI countries appear to be already in a level of human capital and infrastructure development that allows them to extract benefits from governmental online services. In the regressions for the HI countries none of the variables that proxy the

level of e-gov development turned out to be statistically significant. High income countries already exhibit very high levels in the *CCI*, and additional improvements are harder. They also have high levels in the *EGDI/OSI*, and apparently additional improvements in the latter do not influence the *CCI*.

6 Conclusion

Some countries have been successful in achieving prosperity, while some remain poor and shackled by corruption. Corruption undermines investment and innovation, which are essential to growth. Institutional features are important to understand why there are differences among countries and why some of them seem to be unable to overcome this trap. New mechanisms aimed for controlling corruption have been developed worldwide. Digital government tools are nowadays among the most popular ones. They can promote transparency, accountability, and better access to information. Our results clearly indicate that higher levels of e-gov are associated with better corruption outcomes, even when controlling for classical corruption predictors, such as GDP *per capita*, press freedom or bureaucracy. Our results are robust to methods that control for time effects or time invariant country characteristics. However, other sources of endogeneity or omitted variable bias are still a possibility.

A key contribute of our analysis is that the relationship between digital government and corruption varies across income groups. Our results indicate that policy makers from the least developed countries should not look to e-gov as an ultimate solution to fight corruption. First, they should invest in infrastructures and human capital, to guarantee that e-gov services are accessible and the population is literate enough to use them. On the other hand, the upper middle income countries seem to be the ones that can fully exploit the development of e-gov as a key tool for fighting corruption.

The relationship between e-gov and corruption is a hot topic in the e-gov field. With this paper, we show that the investment in ICT tools and their inclusion in governmental bodies could be an effective mechanism to fight corruption. As our measure of e-gov is very broad, future research could focus on different dimensions of electronic government (e.g. open government), to assess which ones are more helpful to improve corruption outcomes. Future research could always focus on other measures of corruption (e.g.: Corruption Perceptions Index) and use instrumental variables estimators in order to control for endogeneity.

Acknowledgments. This paper is a result of the project "SmartEGOV: Harnessing EGOV for Smart Governance (Foundations, methods, Tools)/NORTE-01-0145-FEDER-000037", supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (EFDR). This work was also carried out within the funding with COMPETE reference nº POCI-01-0145-FEDER-006683, with the FCT/MEC's (Fundação para a Ciência e a Tecnologia, I.P.) financial support through national funding and by the ERDF through the Operational Programme on "Competitiveness and Internationalization – COMPETE 2020 under the PT2020 Partnership Agreement."

References

1. Ades, A., Di Tella, R.: Rents, competition, and corruption. Am. Econ. Rev. **89**(4), 982–993 (1999)
2. Andersen, T.: E-Government as an anti-corruption strategy. Inf. Econ. Policy **21**(3), 201–210 (2009)
3. Andersen, K., Henriksen, H., Medaglia, R., Danziger, J., Sannarnes, M., Enemærke, M.: Fads and facts of e-government: a review of impacts of e-government (2003–2009). Int. J. Public Adm. **33**(11), 564–579 (2010)
4. Arendsen, R., Pters, O., ter Hedde, M., van Dijk, J.: Does e-government reduce the administrative burden of businesses? An assessment of business-to-government systems usage in the Netherlands. Gov. Inf. Q. **31**, 160–169 (2014)
5. Belsley, D.: Conditioning Diagnostics: Collinearity and Weak Data in Regression, 1st edn. Wiley, New Jersey (1991)
6. Bertot, J., Jaeger, P., Grimes, J.: 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**(3), 264–271 (2010)
7. Bhatnagar, S.: E-government and access to information. Glob. Corrupt. Rep. **2003**, 24–32 (2003)
8. Bhuiyan, S.: Modernizing Bangladesh public administration through e-governance: benefits and challenges. Gov. Inf. Q. **28**(1), 54–65 (2011)
9. Brunetti, A., Weder, B.: A free press is bad news for corruption. J. Public Econ. **87**(7), 1801–1824 (2003)
10. Cho, Y., Choi, B.: E-government to combat corruption: the case of Seoul metropolitan government. Int. J. Public Adm. **27**(10), 719–735 (2004)
11. DiRienzo, C., Das, J., Cort, K., Burbridge, J.: Corruption and the role of information. J. Int. Bus. Stud. **38**(2), 320–332 (2007)
12. Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A.: The regulation of entry. Q. J. Econ. **117**(1), 1–37 (2002)
13. Estevez, E., Janowski, T.: Electronic governance for sustainable development—conceptual framework and state of research. Gov. Inf. Q. **30**, 94–109 (2013)
14. Faccio, M.: Politically connected firms. Am. Econ. Rev. **96**(1), 369–386 (2006)
15. Grönlund, Å., Horan, T.: Introducing e-gov: history, definitions, and issues. Commun. Assoc. Inf. Syst. **15**(1), 713–729 (2005)
16. Janowski, T.: Digital government evolution: from transformation to contextualization. Gov. Inf. Q. **32**, 221–236 (2015)
17. Kaufmann, D., Kraay, A., Mastruzzi, M.: The worldwide governance indicators: methodology and analytical issues. Hague J. Rule Law **3**(2), 220–246 (2011)
18. Khwaja, A., Mian, A.: Do lenders favor politically connected firms? Rent provision in an emerging financial market. Q. J. Econ. **120**(4), 1371–1411 (2005)
19. Lau, T., Aboulhoson, M., Lin, C., Atkin, D.: Adoption of e-government in three Latin American countries: Argentina. Brazil Mexico. Telecommun. Policy **32**(2), 88–100 (2008)
20. Luna-Reyes, L., Gil-Garcia, J.: Digital government transformation and internet portals: the co-evolution of technology, organizations, and institutions. Gov. Inf. Q. **31**(4), 545–555 (2014)
21. Mauro, P.: Corruption and growth. Q. J. Econ. **110**(3), 681–712 (1995)
22. Ojha, A., Palvia, S., Gupta, M.: A model for impact of e-government on corruption: exploring theoretical foundations. In: Critical Thinking e-Gov, pp. 160–170 (2008)

23. Pathak, R., Singh, G., Belwal, R., Smith, R.: E-governance and corruption-developments and Issues in Ethiopia. Public Org. Rev. **7**(3), 195–208 (2007)
24. Pathak, R., Naz, R., Rahman, M., Smith, R., Nayan Agarwal, K.: E-governance to cut corruption in public service delivery: a case study of Fiji. Int. J. Public Adm. **32**(5), 415–437 (2009)
25. Purian, R.: A decade of e-government indexes: the emerging mode of participative communication. Syst. Signs Actions **8**(1), 43–75 (2014)
26. Shim, D., Eom, T.: E-government and anti-corruption: empirical analysis of international data. Int. J. Public Adm. **31**(3), 298–316 (2008)
27. Shleifer, A., Vishny, R.: Corruption. Q. J. Econ. **108**(3), 599–617 (1993)
28. Schreiber, S.: The Hausman test can be negative even asymptotically. J. Econ. Stat. **228**(4), 394–405 (2008)
29. Schuppan, T.: E-Government in developing countries: experiences from sub-Saharan Africa. Gov. Inf. Q. **26**(1), 118–127 (2009)
30. Svensson, J.: Who must pay bribes and how much? Evidence from a cross section of firms. Q. J. Econ. **118**(1), 207–230 (2003)
31. Treisman, D.: What have we learned about the causes of corruption from ten years of cross-national empirical research? Ann. Rev. Polit. Sci. **10**, 211–244 (2007)
32. United Nations: E-Government Survey 2016: E-Government in Support of Sustainable Development. United Nations, New York (2016)
33. Veiga, L., Janowski, T., Barbosa, L.: Digital government and administrative burden reduction. In: Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance, pp. 323–326. ACM, Montevideo (2016)
34. Veiga, L., Rohman, I.K.: e-Government and the shadow economy: evidence from across the globe. In: Janssen, M., et al. (eds.) EGOV 2017. LNCS, vol. 10428, pp. 105–116. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64677-0_9
35. Wooldridge, J.: Econometric analysis of cross section and panel data, 2nd edn. MIT press, Cambridge (2010)

Local Open Government: Empirical Evidence from Austrian Municipalities

Lisa Schmidthuber[1(✉)], Bernhard Krabina[2], and Dennis Hilgers[1]

[1] Johannes Kepler University, Altenbergerstraße 69, 4040 Linz, Austria
lisa.schmidthuber@jku.at
[2] KDZ – Centre for Public Administration, Guglgasse 13, 1110 Vienna, Austria

Abstract. Local governments have increasingly been applying an open and collaborative approach towards public management during the last years. Accordingly, they aim at increasing accessibility by releasing public data and providing participative decision-making arenas. 'Open government' has also been implemented in Austrian municipalities. This paper takes stock of the current status of open government implementation in Austria by analyzing survey data from city managers. Findings indicate that Austrian municipalities choose releasing public data over involving citizens in decision-making. Although public managers seem to value the principles of an open government, a successful implementation of open government is hampered by resource constraints.

Keywords: Open government · Local government · Public manager

1 Introduction

Decreasing levels of public trust, low citizen satisfaction with the political-administrative system, and a changing political landscape are challenging government institutions, public sector organizations, and political parties alike. These acceptance and legitimacy problems assert pressure on the political-administrative system to change and innovate the present process of public service delivery and policy-making. In fact, various public institutions respond to these challenges by providing citizens free access to public data and possibilities of participative decision-making.

Various scholars (e.g. [13, 16, 17, 20]) gave promising results of 'open innovation' as used in the public sector. They suggested that online platforms, social media, and mobile applications are leveraged to enhance the dialogue between government and citizens. Knowledge flows in digital platforms further hold great potential for value-creation both for society and participants [16]. The implementation of open government was, however, only investigated by pointing to specific open government projects or best practices so far (e.g. [20]). Unfortunately, research has not yet examined to what extent open government is implemented across local governments, independent from size. Consequently, one cannot formulate more general statements on factors

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 110–119, 2018.
https://doi.org/10.1007/978-3-319-98690-6_10

facilitating or hampering open government implementation. Moreover, there is little research on city managers' perceptions on open government [4, 25].

This paper thus contributes to the research on open government by taking stock of open government implementation in Austrian municipalities. It intends to evaluate the current status of adoption, analyze municipalities' capability for implementation, and shed light on city managers' attitudes towards governmental openness.

The remainder of this paper is structured as follows: In Sect. 2, we introduce the concept of open government. Section 3 describes the data and method used in this study. In Sect. 4, empirical findings are presented. Section 5 encompasses a discussion of our findings, gives implications of this research, and points to limitations and avenues for further research.

2 Open Government

Open government as a 'multilateral, political, and social process' [23, p. 2] is widely understood as a concept that transforms governments and is meant to connect government with citizens who are supposed to introduce innovative solutions based on their local knowledge and experience [18]. Meijer, Curtin, and Hillebrandt [15] associate open government with citizens' vision of 'what is going on inside government' and their voice in interactive terms, the term denotes governmental activities for the benefit of transparency, participative decision-making, and collaborative activities between policymakers and citizens (e.g. [3, 12]). An 'open government' thus encompasses an increased level of openness in terms of information and decision-making and can be seen as the comprehensive redesign of politics and administrative activities according to the principles of modern public management and public governance (see Fig. 1).

The public sector utilizes various ways to promote transparency of government action, accessibility of government services and information as well as the integration of externals by leveraging modern information and communication technology [11, 13, 15–17, 23]. Significantly, opening government to the public sphere requires a border crossing for the respective public organizations, meaning that new forms of cooperation are implemented and external sources of knowledge are consulted for specific issues [8]. More typical forms of participation such as elections and referendums fail to adequately transfer either knowledge or needs from society to government. Openness and free access to information, however, facilitate the interaction between government and civil society and promote a participatory government [14]. Citizens and experts provide know-how that governments then apply in problem solving and draw on when developing innovative strategies and policies [13, 21].

With this in mind, integrating citizens into organizational processes and governmental decision may result in various benefits. First, applying open government is assumed to enhance government-citizen relation. Opening up government improves perceptions of fairness among the public, strengthens democracy insofar as personal responsibility and public spirit can be enhanced, democratic accountability, and trust in government [7, 10]. Second, openness in the public sector may lead to advantages for public administration and its organizational processes, as the new approach benefits

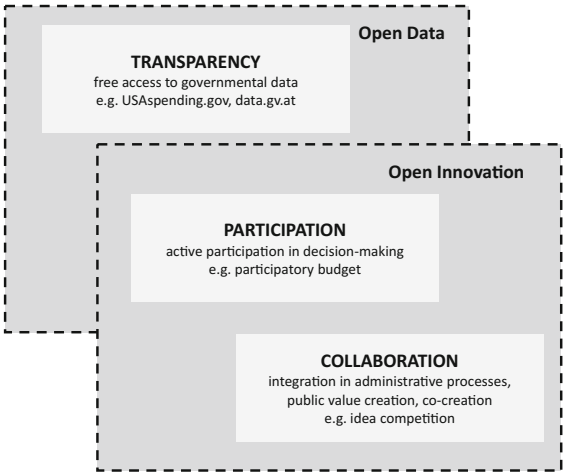

Fig. 1. Open government [20]

organizational and policy performance [6, 18], efficiency and effectiveness of public services and governmental procedures [16, 18, 23] as well as quality enhancements [9, 16, 19]. In summary, open government activities are assumed to enhance public value.

3 Data and Methods

To explore the current status of open government implementation in Austria, we collected survey data in Austrian municipalities. Austria is a federal state consisting of nine regions and 2,100 local governments[1]. Municipalities are characterized by local autonomy, which means that they are governed by a local council with the mayor as its head. The mayor is elected by the local council or local inhabitants for five or six years. Whereas the mayor represents the political decision-maker of local government, the city manager or chief executive officer is subject to directives from the mayor and responsible for operational matters. In November 2016, city managers of all Austrian municipalities received an online questionnaire with a number of closed questions. 235 city managers completed the questionnaire.

4 Findings

4.1 Sample Description

Table 1 summarizes key characteristics of the sample population, and contrasts it from the Austrian population. Expect for Vienna, individuals from all federal states respond to our survey invitation. The greatest share of respondents are from Upper Austria as

[1] At the time of the analysis, Austria had 2,100 municipalities.

the second largest federal state. Furthermore, the sample varies among municipal size. About one third of all respondents work in small municipalities, one third in large municipalities, and one third is in between. In addition, the level of debt of sample municipalities is comparable to the actual Austrian distribution. Consequently, we conclude that the survey sample resembles the Austrian structure of municipalities in terms of federal state, municipal size, and level of debt.

Table 1. Sample description

	Sample distribution	Austrian distribution
Federal state		
Burgenland	12 (5.11%)	171 (8.14%)
Carinthia	20 (8.51%)	132 (6.29%)
Lower Austria	58 (24.68%)	573 (27.29%)
Upper Austria	80 (34.04%)	442 (21.05%)
Salzburg	15 (6.38%)	119 (5.67%)
Styria	20 (8.51%)	287 (13.67%)
Tyrol	21 (8.94%)	279 (13.29%)
Vorarlberg	9 (3.83%)	96 (4.57%)
Vienna	–	1
Municipal size		
Small	80 (34.04%)	830 (39.52%)
Intermediate	91 (38.72%)	721 (34.33%)
Large	64 (27.23%)	549 (26.14%)
Level of debt per capita		
Low	144 (61.28%)	1364 (64.95%)
Intermediate	59 (25.11%)	409 (19.70%)
High	32 (13.62%)	303 (14.60%)
Observations	**235**	**2,100**

Source: Statistics Austria

4.2 Open Government Implementation in Austrian Municipalities

In order to evaluate the status of open government implementation in Austria, we developed a list with practices associated with open government. Figure 2 illustrates the survey results and indicates if the responding municipality has already implemented a practice, is intended to implement it, or is not willing to take action.

First, findings show that the status of implementation greatly varies across municipalities. For example, 63 municipalities use social media channels to communicate with citizens, whereas 132 municipalities have no intention to set up a social media account. Second, the results indicate a great difference in the implementation status in terms of type of practice. On the one hand, the majority of city managers report to exchange with other municipalities, release public data on open data portal,

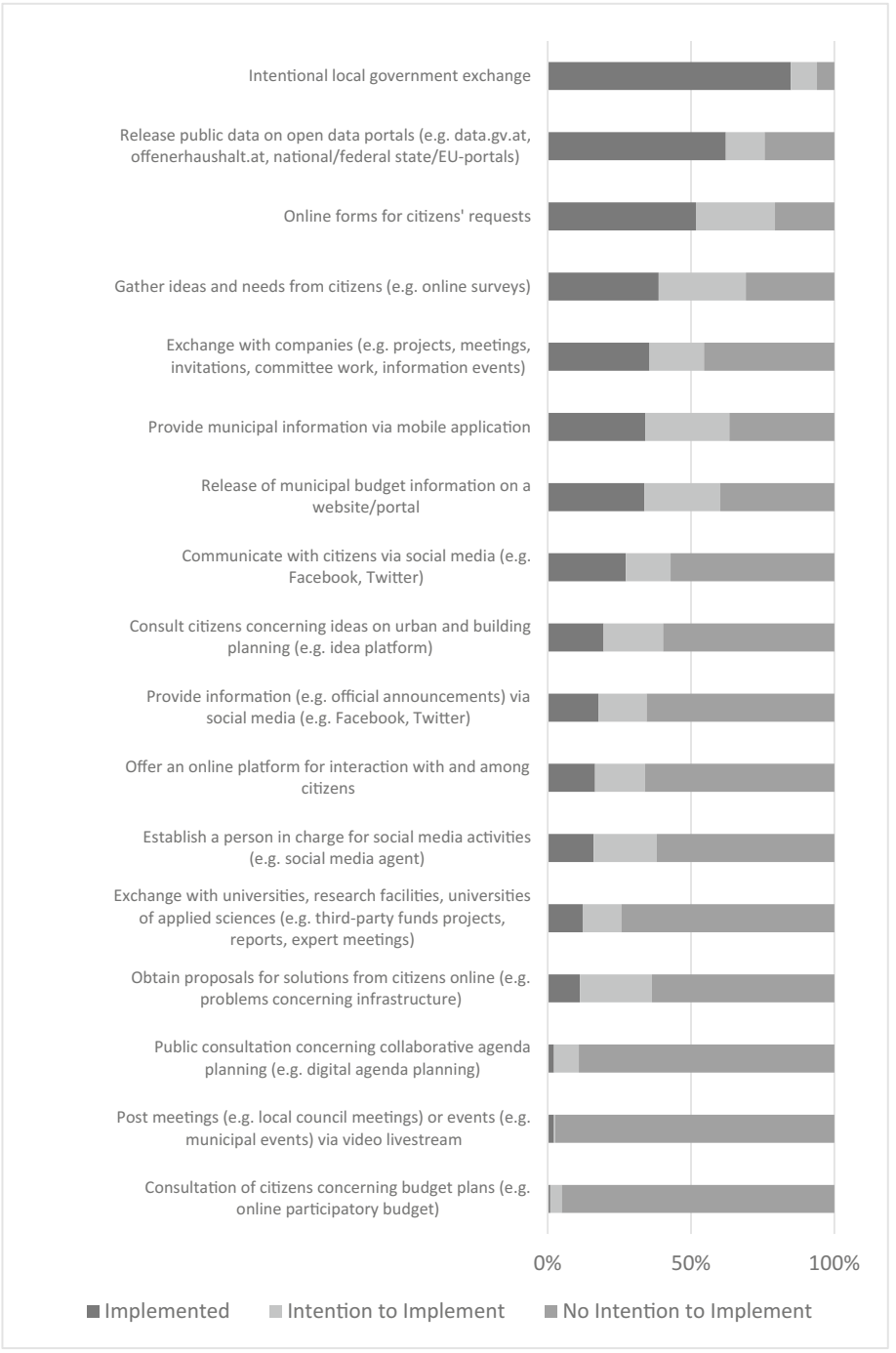

Fig. 2. Open government practices in Austria.

and provide online forms for citizens' requests. On the other hand, however, only very few municipalities consult citizens in terms of budgeting or agenda planning and rather have no intention to do so in the future.

4.3 Capability to Implement Open Government

Based on the main ideas of the resource-based view, more resources are associated with a higher level of organizational performance [22]. Organizational resources are also shown to influence public innovation [2]. Open government implementation thus requires organization's capability to provide relevant practices. Accordingly, municipalities must have resources to ensure transparency, provide citizen participation, and foster collaboration. We thus ask city managers if current resources are adequate for implementing open government practices. As illustrated in Fig. 3, municipalities seem to suffer from resource scarcity. Whereas technical equipment for disclosing data and integrating citizens as well as security arrangements are available to a certain level, financial and human resources do not hold for adopting open government.

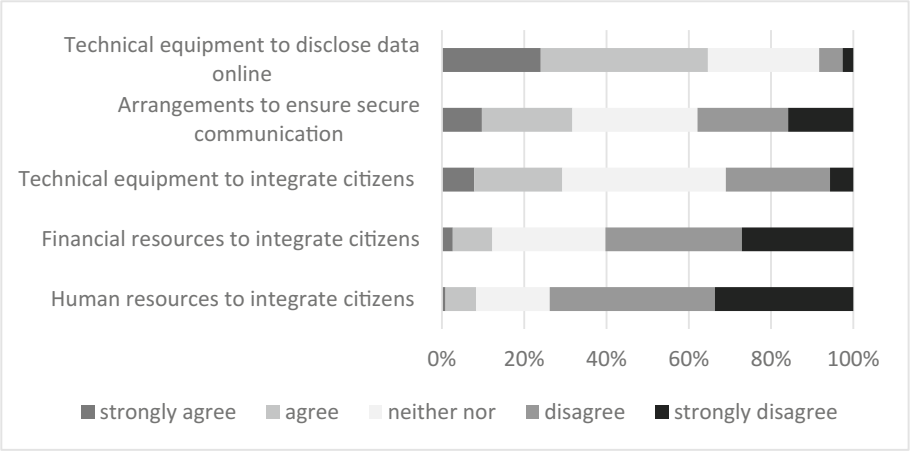

Fig. 3. Resources. (N = 227–229)

Next to material resources, open government implementation requires task knowledge. Accordingly, public employees have to know how to implement open government. In Fig. 4, we give an overview on respondents' statements on the level of task knowledge. The results show that the majority of municipalities have knowledge in releasing public data. Furthermore, they seem to know how to transfer citizens' input into improvements of organizational activities. In addition, about 60% of municipalities are able to manage citizens' contributions.

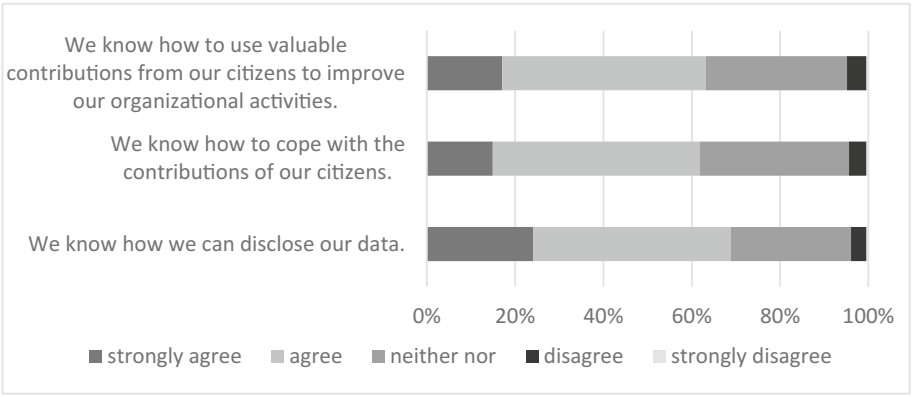

Fig. 4. Knowledge. (N = 228–229)

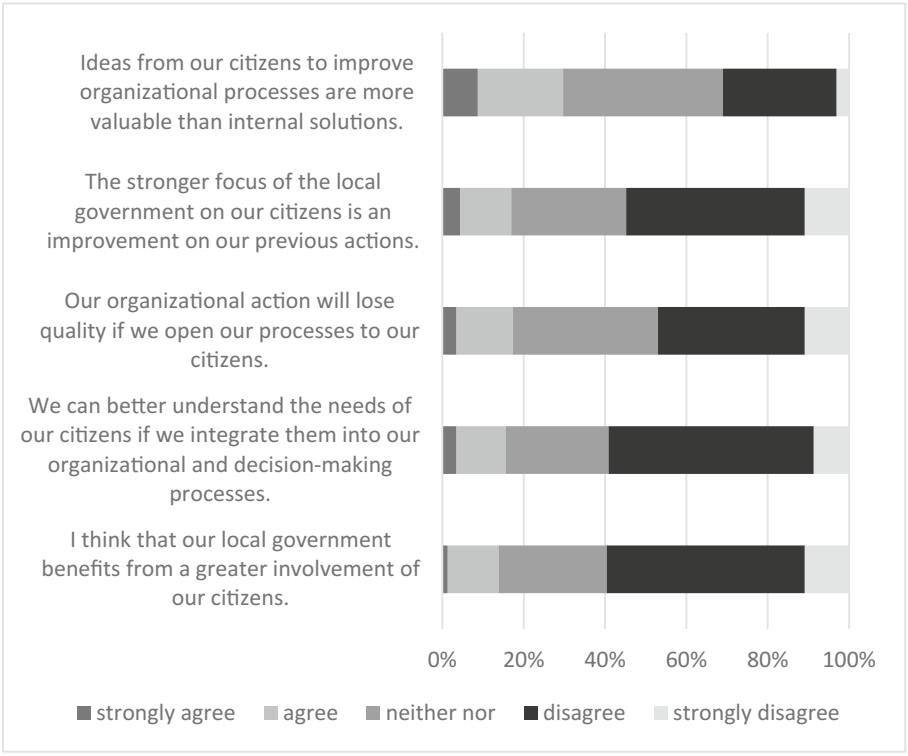

Fig. 5. Attitudes towards open government. (N = 229–230)

4.4 Attitudes Towards Open Government

Next to organizational factors crucial for open government adoption, we concentrate on decision-makers' attitudes towards innovative practices. Shedding light on the

perceptions of municipal city managers is important, as they are main decision-makers in implementing public change and innovation [5, 24], and thus influence strategic decisions [1]. Accordingly, we first focuses on city managers' attitudes towards government. Results in Fig. 5 show that about 30% of city managers seem to value the ideas from citizens, even more than internal ones. However, only few managers agree that citizen involvement can lead to an improvement of organizational action.

Second, we question the perceived value of open government. As summarized in Fig. 6, city managers seem to value open government practices in terms of citizen integration and transparency. In contrast, they are less convinced of open government as a tool to compensate losses in public trust and citizen confidence.

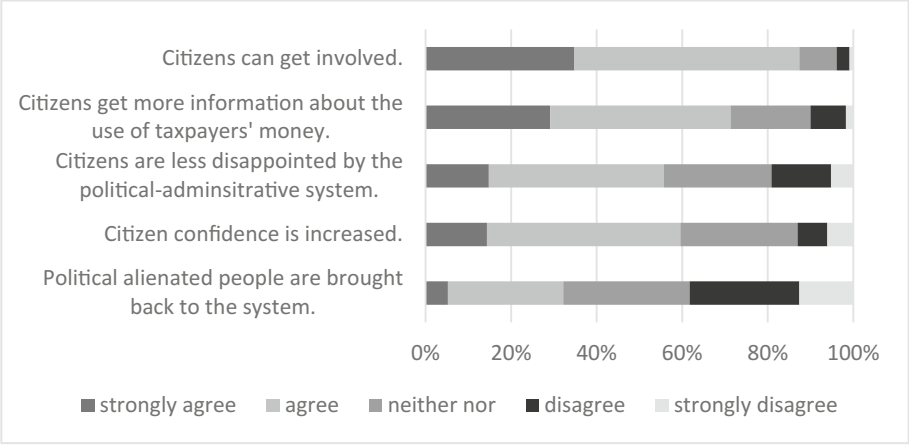

Fig. 6. Outcomes of open government. (N = 230)

5 Discussion and Conclusion

The findings of this study show a diverse picture of open government in Austrian municipalities. First, while municipalities value the exchange of experiences with each other, they seem to be quite reluctant to adopt open government ideas in terms of participation and collaboration. Second, city mangers seem to rate the outcomes of open government rather positively. However, they refer to a lack of available organizational and financial resources as a barrier to implement open government in their organizations successfully. One could thus explain a low level of open government adoption with a lack of resources. While scarce resources are certainly a problem in some municipalities, however, the question of re-allocation resources from other tasks to the tasks at hand regarding the open government agenda might arise. Here, the question of attitude towards open government speaks a clear language: when almost 60% of public managers do not think that their local governments can benefit from a greater involvement of their citizens, why should they shift resources to topics they are not ordered to do so by the politicians in charge or required by law.

Concluding, this study confirms two observations in Austria. First, the open government discussion in Austria has been focusing mainly on open government data (OGD), rather than participative decision-making and collaboration. Whereas municipalities are ready to disclosure data, they seem to be not willing to involve external actors in organizational processes. For example, whereas many German cities have already adopted various rounds of participatory budgeting, Austrian municipalities are not taking up on this practice. Second, survey findings indicate that the open government agenda is valued in general. Few Austrian cities have already initiated large open government projects (see [20]). However, the majority of Austrian municipalities suffer from resource constrains for realizing the open government agenda. Consequently, open government has to be move to political representatives who has to actively push and promote the topic.

Further research has thus to shed light on the role of politicians and investigate their perception toward open government. A further avenue of research relates to the extent to which municipalities have adopted open government. Whereas this study makes a first step towards illustrating what open government means in practice, further research is recommended to elaborate a scale and give a more holistic picture of the quantity and quality of open government practices. Furthermore, we know little about what public datasets are released, how often they are downloaded by citizens, and if citizen expectations are fulfilled.

References

1. Damanpour, F., Schneider, M.: Phases of the adoption of innovation in organizations: effects of environment, organization and top managers. Br. J. Manag. **17**(3), 215–236 (2006)
2. De Vries, H., Bekkers, V., Tummers, L.: Innovation in the public sector: a systematic review and future research agenda. Public Adm. **94**(1), 146–166 (2016)
3. Evans, A.M., Campos, A.: Open government initiatives: challenges of citizen participation. J. Policy Anal. Manag. **32**(1), 172–185 (2013)
4. Ganapati, S., Reddick, C.G.: The use of ICT for open government in US municipalities: perceptions of chief administrative officers. Public Perform. Manag. Rev. **37**(3), 365–387 (2014)
5. Hansen, M.B.: Antecedents of organizational innovation: the diffusion of new public management into Danish local government. Public Adm. **89**(2), 285–306 (2011)
6. Harrison, T.M., et al.: Open government and e-government: democratic challenges from a public value perspective. Inf. Polity **17**(2), 83–97 (2012)
7. Herian, M.N., Hamm, J.A., Tomkins, A.J., Pytlik Zillig, L.M.: Public participation, procedur-al fairness, and evaluations of local governance: the moderating role of uncertainty. J. Public Adm. Res. Theory **22**(4), 815–840 (2012)
8. Hilgers, D.: Open government. Zeitschrift für Betriebswirtschaft **82**(6), 631–660 (2012)
9. Hilgers, D., Ihl, C.: Citizensourcing: applying the concept of open innovation to the public sector. Int. J. Public Particip. **4**(1), 67–88 (2010)
10. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. **29**(4), 258–268 (2012)
11. Krabina, B., Lutz, B.: Open Government Implementation Model (2016). http://www.kdz.eu/en/open-government-implementation-model. Accessed 15 Mar 2018

12. Lee, G., Kwak, Y.H.: An open government maturity model for social media-based public engagement. Gov. Inf. Q. **29**(4), 492–503 (2012)
13. Linders, D.: From e-government to we-government: defining a typology for citizen coproduction in the age of social media. Gov. Inf. Q. **29**(4), 446–454 (2012)
14. Meijer, A.: Government transparency in historical perspective: From the ancient regime to open data in the Netherlands. Int. J. Public Adm. **38**(3), 189–199 (2015)
15. Meijer, A.J., Curtin, D., Hillebrandt, M.: Open government: connecting vision and voice. Int. Rev. Adm. Sci. **78**(1), 10–29 (2012)
16. Mergel, I.: Opening government: designing open innovation processes to collaborate with external problem solvers. Soc. Sci. Comput. Rev. **33**(5), 599–612 (2015)
17. Nam, T.: Suggesting frameworks of citizen-sourcing via Government 2.0. Gov. Inf. Q. **29** (1), 12–20 (2012)
18. Neshkova, M.I., Guo, H.: Public participation and organizational performance: evidence from state agencies. J. Public Adm. Res. Theor. **22**(2), 267–288 (2011)
19. Salge, T.O., Vera, A.: Benefiting from public sector innovation: the moderating role of customer and learning orientation. Public Adm. Rev. **72**(4), 550–560 (2012)
20. Schmidthuber, L., Hilgers, D.: Unleashing innovation beyond organizational boundaries: exploring citizensourcing projects. Int. J. Public Adm. **41**(4), 268–283 (2018)
21. Schmidthuber, L., Hilgers, D., Gegenhuber, T., Etzelstorfer, S.: The emergence of local open government: determinants of citizen participation in online service reporting. Gov. Inf. Q. **34** (3), 457–469 (2017)
22. Williamson, O.E.: Strategy research: governance and competence perspectives. Strat. Manag. J. **29**(12), 1087–1108 (1999)
23. Wirtz, B.W., Birkmeyer, S.: Open government: origin, development, and conceptual perspectives. Int. J. Public Adm. **38**(5), 381–396 (2015)
24. Yukl, G.: Leadership in Organizations, 6th edn. Pearson Prentice Hall, Upper Saddle River (2006)
25. Zhang, F., Feeney, M.K.: Managerial ambivalence and electronic civic engagement: the role of public manager beliefs and perceived needs. Public Adm. Rev. **78**(1), 58–70 (2018)

Who Is Measuring What and How in EGOV Domain?

Joana Carvalho[1(✉)] and Delfina Soares[2]

[1] University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal
joanacarvalho87@gmail.com
[2] United Nations University, Campus de Couros, 4810-445 Guimarães, Portugal
soares@unu.edu

Abstract. Governments around the world have accepted information and communication technology as a facilitator to reform, transform and modernize the governance activity - EGOV. Although EGOV is accepted to bear several benefits, government accountability demands its measurement and assessment for management and improvement. Many tools and instruments have been proposed by researchers for measuring, assessing and monitoring different aspects of EGOV, but their value has not been availed by researchers and practitioners alike. This happens because evaluation tools are dispersed among various sources and there is also no systematized framework to support the analysis and selection of the adequate tool for specific situations. This work intends to answer these issues by characterizing the literature available in the context of EGOV measurement, assessment and monitoring to generate a strong base of knowledge oriented towards the creation of a future catalog of tools and instruments for EGOV evaluation, and to present a conceptual framework to support the choice of an adequate tool from such catalogue.

Keywords: Electronic governance · E-governance · EGOV · Evaluation
Measurement · Assessment · Monitoring · Literature review

1 Introduction

Information and Communication Technology (ICT) is nowadays commonly used to support the execution of States' multiple governance activities in its different aspects – Politics, Administration and Society. Each of these aspects, and the relationships between them, are fundamental spaces where ICT can be used as a facilitator and catalyzer to promote the reform, transformation and modernization of the overall State governance activity [1, 2]. This use of ICT, and particularly the Internet, as a tool to achieve better governance, is what is currently called electronic governance (EGOV).

EGOV is advocated to offer a set of benefits, contributing to the existence of more efficient, effective and transparent public institutions, to improve services delivery, and for a more participatory and engaged society [3]. EGOV benefits, although commonly accepted by policy-makers [4], must be measured and assessed to ensure government accountability [5–7] and to support state functions and services [8]. Measuring and assessing are also relevant activities to analyze the state of EGOV development as well

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 120–131, 2018.
https://doi.org/10.1007/978-3-319-98690-6_11

as to inform strategy and policy formulation [9]. For this, governments have been measuring, assessing and evaluating their EGOV initiatives [10] following the premise of "measuring for management and improvement" [11], guided by questions such as "are we doing it right?" and "what are others doing?".

Many tools and instruments for measuring, assessing and monitoring different aspects of EGOV have also been proposed by researchers and can be found in the literature. While these may be very useful tools, their value may not have been fully exploited either by researchers or, particularly, by practitioners, for two main reasons. First, because the many existing tools are spread through a large body of sources, being difficult to find. At the best of our knowledge, there is no catalog or repository of EGOV evaluation tools that provides researchers and practitioners with a holistic view of the overall existent set of tools. Second, there is also no systematized conceptual framework to support the analysis and selection of and adequate tool to be used in specific assessment situations.

This work aims at contributing to mitigate these problems by answering a very preliminary question: "who is measuring what and how in the EGOV domain?" and providing a literature overview for researchers that resumes the state of the art regarding EGOV measurement and evaluation.

By replying to this question this paper will achieve two main objectives: (i) to characterize the research that has been conducted in the context of EGOV measurement, assessment and monitoring area, thus creating a rich and sound EGOV assessment base of knowledge fundamental for the creation of a future catalog, and (ii) to present a conceptual framework to support the analysis about the adequacy of a tool, and the selection process of such suitable tool.

The remainder of the paper is structured as follows: Sect. 2 highlights the importance and complexity of EGOV evaluation. The study design is presented in Sect. 3 and the results of the study conducted are reflected upon in Sect. 4. A first version of the conceptual model towards the selection of the appropriate instrument of e-government evaluation is presented in Sect. 5 and conclusions and future work are included in Sect. 6.

2 The Importance and Complexity of EGOV Evaluation

EGOV evaluation is a relevant topic not only for government agencies but also for other stakeholders who take an interest in the area. According to [12] there are different groups who take interest in measuring, assessing and monitoring EGOV: large international organizations that deal with e-government measurement globally; global independent organizations; multinational consulting companies; academic institutions and its non-profit research centres; national institutions or national associations for ICT in the public sector in a single country or region; and researchers groups.

Many of these provide international benchmarking studies while others focus on a single country, region or municipality. For any case, evaluation results are intended to: justify the investment made [13], justify a projects' worthiness, and provide some kind of learning [14]. Specifically for EGOV assessment, we can ascertain that evaluation is vital to understand the level of EGOV development, review objectives, strategies and

action plans, discover strengths and weaknesses, define new guidelines, search for best practices and compare organisations at the various levels [15, 16]. Meeting national strategy goals is vital for governments and therefore, effective, efficient and quality evaluation activities are required [17] to avoid loss of control that will consequently lead to loss of resources and failure to accomplish such goals [18].

Measuring and assessing EGOV is not, however, a simple task. The lack of unified, holistic perspectives to EGOV evaluation leaves practitioners with the strategic decision of what should they be focusing and how can adequate measurements be formulated [19].

We know that ICT is constantly evolving and a continuous adaptation to its growth implies a continuous evolution of the way e-government is assessed. Besides, e-government projects do not usually have immediate results and take a long time to permeate [20], making it difficult to demonstrate value and, consequently, hampering EGOV development [16]. These issues lead to a disperse set of evaluation goals that as [10] states, results in 'an eclectic mixture of exercises undertaken in different ways for different purposes at different times by different people and with different audiences in mind'. This is a clear picture of the state of the art when it comes to measuring and monitoring EGOV.

Evaluation is a "job for the brave" because the hard work required and its complexity can be extenuating [18]. Especially in the EGOV context, many challenges arise [1]. A robust evaluation would enable the comparison of benefits that can be distinguished as direct and indirect [14]. Direct benefits are measurable or quantifiable while indirect benefits are more qualitative and not so easily measured, such as organizational, social, political or cultural aspects [13]. Benefits vary according to the initiatives goals and objectives and their measurement also varies according to the stakeholder perspective. This many times results in too simplistic evaluations focused on what is easy to measure [21], like the front-office, the visible side of e-government, ignoring back-office reorganization that could improve service efficiency [22].

Two other essential complexities regarding EGOV evaluation are appointed by [13]. The first one focuses on the multiple perspectives involved. Politicians, policy-makers, e-government development leaders, citizens, etc., are some of the perspectives that can be included. Answering the needs of every perspective can be a daunting task, not only because they are so different but they are also conflicting [16]. The second aspect is the social and technical context of use. Public sector cannot rely solely on economic values for it has a clear responsibility towards citizens and society in providing equality, openness and transparency values, among others that make e-government evaluation much more a social science [16].

Other complexities pointed in the literature include: the lack of a clearly defined purpose in some initiatives about what should be compared and measured, making it difficult to be adapted to specific national or regional contexts and priorities [22], and the lack of a comprehensive and holistic assessment [21, 22], although such an assessment would entail huge funding, time and other resources which are not usually available in public administration [1, 23].

3 Study Design

This work is the result of a meta-analysis of the published literature regarding EGOV evaluation. This ended in an extensive literature review intending to understand what the academic community is doing within this subject.

Articles were selected from Scopus in December 2016. Scopus was the only bibliographic database selected because it is one of the databases with most recognition. Articles selection was achieved using a word combination of 'EGOV', 'egovernment' or 'e-government' with evaluation, measurement, assessment and monitoring, with no date limitation. A total of 2428 references were identified. After retrieving complete information on each article, the process of reference filtering described in Table 1 was performed to achieve the final number of articles to be analysed.

Table 1. Process of references downsizing

Activity	References left
Removal of duplicate references	1762
Removal of general references of conferences and books	1587
Removal of non-relevant references by abstract reading	581
Removal similar references from the same author	529
Removal of articles not fully available on online sources	472
Removal of articles in foreign language (unknown for the researchers)	456
Eliminating retracted articles	454

Table 2. Categories for literature analysis

Category	Description
Year	Refers to the publication year of the article
Object	Refers to the object of study, focus of evaluation or measurement described by the author(s)
Where	Refers to the country where the research (when practical application happened) took place
Perspective	Refers to the perspective used when applying the instrument. It can be: end user, service provider or a combination of these. If the instrument measures results from user interviews, surveys, website consultation, amongst others, it was considered as "end user", since it provides citizen's perceptions. If elements for evaluation are provided solely by the agency where the evaluation takes place it was considered "service provider". Attention is needed for situations where the participants in the evaluation are government employees but they are considered the end users of the system. In those cases perspective is considered as "end user"
Government level	Refers to the level of government considered. It can be: local, regional, or central (national, federal), or any combination of these when more than one level is assessed
Has indicators	This field was marked if the article provided indicators and metrics to measure the object of evaluation
Affiliation	Refers to the institution/organization the author(s) is/are from

The remaining 454 articles were analysed and categorized using a table created for the literature review process encompassing several categories, as described in Table 2.

In some categories the classification NS (Not Specified) or NA (Not Applied) are applied regarding articles where the authors did not specify such category or where the category is not applied, respectively.

The set of categories defined in Table 2, was applied to each of the 454 papers, providing a huge base for understanding the current state of the art in EGOV evaluation. The main findings from this analysis are presented in the following section.

4 EGOV Evaluation Literature

This section provides a characterization of the research that has been conducted in the EGOV evaluation area.

EGOV evaluation literature starts in the early 2000s although e-government research begins in the mid 1990s [24]. A first look to the publishing years (Fig. 1) reveals a growing trend in the area beginning in 2002 until 2011, where the highest number of publications occurs. From 2011 onward, the number of publications has decreased but has been stabilizing. This evolution follows the typical evolution path exhibited by most research themes, which usually start with a low number of publications that grow slowly, having then an exponential rising until reaching a peak of researchers interest, after which suffer a slight decrease and tend to stabilize.

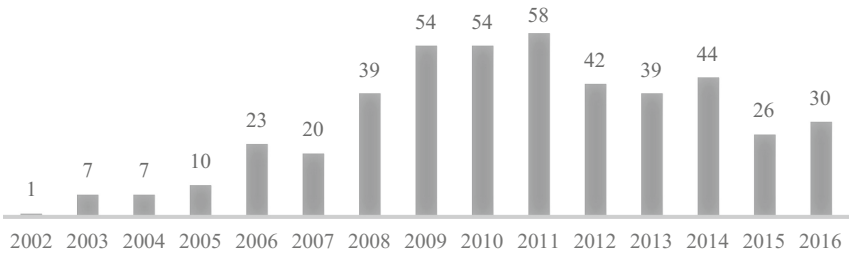

Fig. 1. Papers distribution by publication year

4.1 Who Is Conducting EGOV Evaluation

Presented research is mostly provided by the academic community which comprises over 80% of the selection (Table 3). This is, however, not surprising since the sources used for this analysis were retrieved from Scopus.

Some research results from partnerships between private companies, government agencies, research institutes, or national, regional and international organizations. The most common association is between Academic Institutions and Government Agencies followed by Academic Institutions with Private Companies or Research Institutes.

Table 3. Authors affiliation

Author's affiliation	Total of articles (%)
Academic Institutions	85.5%
Government Agencies	2.4%
Research Institutes	1.5%
National or International Organizations	1.3%
Private Companies	0.4%
Partnerships	8.6%

Most of the selected authors contributed with a single research publication and few have two or more publications leading us to conclude that not many authors follow a research line within the EGOV evaluation area.

Academic authors that mentioned their full affiliation come essentially from Computer Science, Information Science, and Information Systems departments or schools. Other affiliation areas, but with less expression in our sample, are Management, Business Administration, Business Information, Economic Sciences, and Political Sciences.

4.2 What Is Being Evaluated

The object of evaluation – i.e. the focus of evaluation or measurement – is not always well described by researchers and sometimes distinct descriptions appear for the same object of evaluation. Even so, if, for example, several authors mentioned they were measuring 'quality' but definitions were different, we assumed 'quality' to be the object of evaluation for all of them.

Table 4 presents the list of different objects of evaluation that were found in the 454 papers analysed. The objects of evaluation listed are those that were found in at least 2

Table 4. Object of evaluation

Object of evaluation	No. of articles	Object of evaluation	No. of articles
Website evaluation	121	Information systems	5
e-Service	53	Open government data and initiatives	4
Performance	34	Risk assessment	4
e-Readiness	28	Policies	3
Maturity levels	18	Benefits	3
User satisfaction	18	Administrative burden	3
Impact	17	e-Participation	3
Benchmarking	17	ICT investment	3
Security	10	Public value	3
Success	7	Value	3
e-Health	6	Capability	2
m-Government	5	Other	81
Interoperability	5		

papers. The category *Other* includes various objects of evaluation mentioned in the literature that only come up once in the set of papers analyzed.

As [27] states, measuring EGOV has traditionally been focused on measuring and benchmarking websites and their use. Looking at Table 4 we see that there is indeed a big gap between *Website Evaluations* and other focus of study.

Looking particularly to the studies focused on the *Website Evaluation* object (Fig. 2), *Quality* of websites or portals was the specific aspect most measured and assessed by researchers, followed by websites *Accessibility*. *Usability*, *Performance* and *User Satisfaction* are also specific aspects of *Website Evaluation* that have been assessed. The *General* category includes articles that reported a *Website Evaluation* study but with no regard to any specific aspect. The *Other* category includes all the specific aspects of *Website Evaluation* that have been the focus of assessment in only one study.

A curious aspect about *Website Evaluation* is the down peak of publications that it suffers in 2011, just when the selected literature reaches its peak of more publishing.

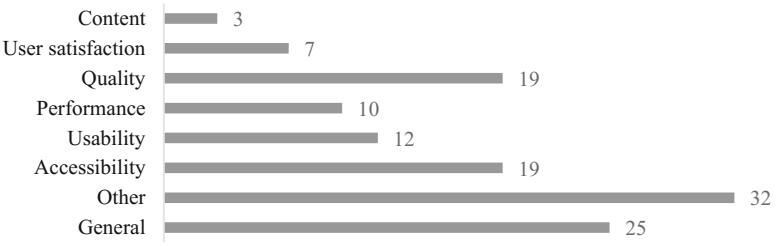

Fig. 2. Specific focuses of *Website Evaluation*

Regarding the geographical distribution of the assessment studies, a distribution by continent is presented in Fig. 3, and over 100 different countries were identified in the studies. Asia and Europe are the two best represented continents with the Americas following just behind. China is clearly the country where most of the selected research takes place with 54 publications, followed by the USA with 24 and some European countries with around 10 publications each.

Twenty-nine authors found it relevant to make a distinction throughout their articles between developed and developing countries, based on the premise that EGOV initiatives are different for these types of countries when it comes to developing strategy, implementation, and utilization [13]. This work maintains this distinction based on the United Nations classification of developed and developing economies presented in the World Economic Situation and Prospects 2017 report. Our ideal is that if such differences exist, evaluation of such initiatives also encompasses social, cultural, technical and political differences and therefore it is important to make such distinction. Through the literature review we found 175 researches located in developing countries and 134 in developed countries.

Looking at the government level in which measurement instruments are put into practice, central and local levels are the most recurrent ones (Fig. 4), although literature

suggests not enough attention is given to local level where citizens are more in contact with government and feel the effect of e-government initiatives more closely [25].

Some authors mention that their evaluation method, tool or framework can be used at any government level (All).

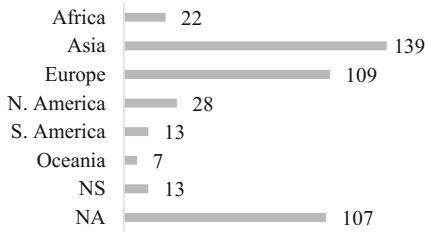

Fig. 3. Selected literature distribution by continent where research took place

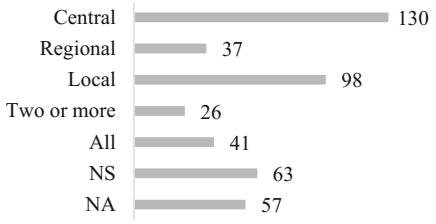

Fig. 4. Level of government distribution for research application

Local level evaluation has been a trend more recently (2015 and 2016), as it was in some of the earlier years (2006 and 2007), while central level evaluation is almost always continuously the focus of the evaluation research (Fig. 5).

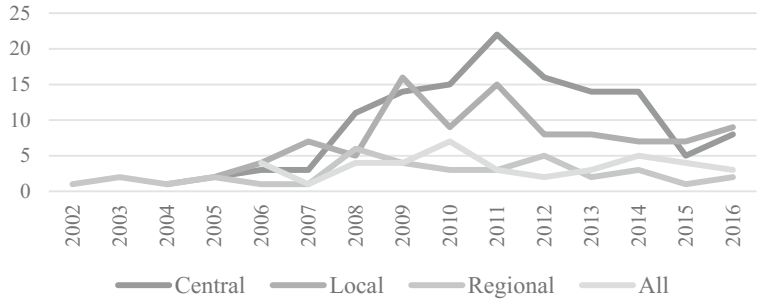

Fig. 5. Level of government described in the publications - distribution by year

Different levels of government have different scopes, objectives and constraints [26] and it is therefore relevant to find and analyze differences between instruments applicable to each of the levels considering what is being assessed for each level.

4.3 How Is Evaluation Being Conducted

Articles were classified according to the type of contribution they provide for the EGOV evaluation area. Four types were considered: *Theory*, if the presented work was theoretical; *Instrument/Framework Design* if an instrument, framework, model or tool design/development was described in the article; *Instrument/Framework Application* if

it describes the application of an instrument, framework, model or tool in a practical case; and *Other* if none of the latter was appropriate. In many cases (Table 5) *Instrument/Framework Design and Application* were both selected.

Table 5. Type of article

Type of article	No. of articles
Theory	64
Instrument/framework design	107
Instrument/framework application	48
Instrument/framework design and application	220
Other	21

In the cases where an instrument/framework was designed or applied, the perspective of the study was analyzed. Two main perspectives were considered: *end user*, when the evaluation is based on information provided by end user, gathered through user interviews, surveys, website consultation, amongst others (demand-side perspective), and *service provider*, when the evaluation is based on information provided by government agencies or entities (supply-side perspective). As noted, there are situations where the participants in the evaluation are government employees but they are considered the end users of the system. In those cases, the perspective considered was "end user".

Figure 6 shows the perspective distribution with end user being the most used perspective to perform e-government evaluations. Some works combine both perspectives.

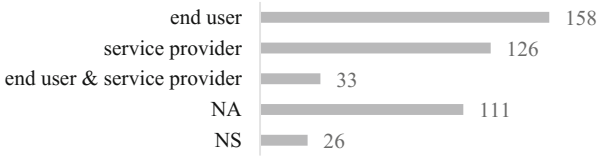

Fig. 6. Number of publications by perspective used in the literature

5 Conceptual Framework for EGOV Evaluation Instrument Characterization

During the literature review, a set of relevant concepts related to EGOV evaluation instruments have been identified. These concepts are aligned with the categories presented in Table 2 and represent useful dimensions that should be considered when analysing, comparing and selecting evaluation instruments. These concepts are organized in the conceptual framework depicted in Fig. 7.

As shown, there are three main dimensions for characterization of an instrument: *object of analysis*, *perspective* and *context*.

Each evaluation instrument has (and must focus on) a specific *object of analysis* that constitutes the focus of the evaluation. This focus of analysis may be studied in different contexts.

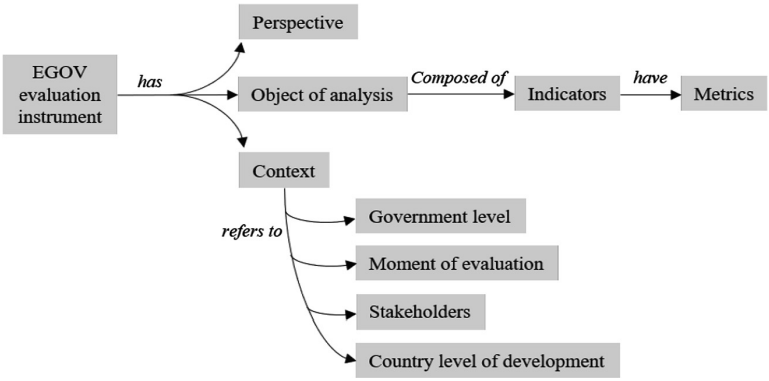

Fig. 7. Conceptual framework for EGOV evaluation instrument characterization

It is thus relevant to characterize and define the *context* of an instrument in what concerns either (i) the level of government considered (local, regional, central (national, federal)), (ii) the level of development of the country where the evaluation is conducted (developed and developing countries), (iii) the stakeholders to whom the evaluation is relevant, and (iv) the moment when the evaluation occurs. It is also fundamental to characterize and define the *perspective* (end user or service provider) that the evaluation considers.

6 Conclusion and Future Work

This work intended to achieve two main goals: (i) to characterize the research that has been conducted in the context of EGOV measurement, assessment and monitoring area and (ii) to present a conceptual framework that would support the analysis and selection process of the most adequate tool for a specific evaluation situation.

The relevance and complexity of EGOV measurement, assessment and monitoring for governments and practitioners is patent in the literature available about the subject and was briefly presented in this article. Literature review of 454 articles retrieved from Scopus gave us a general picture of who has been studying the topic, what are the objects of evaluation, at what level of government are tools and instruments applied, and how is evaluation being conducted regarding the types of publications and the perspective chosen for such evaluations. Our analysis provided enough basis to argue that measurement tools and instruments are indeed dispersed and numerous, and as so, a catalogue of such initiatives could be helpful for researchers and practitioners alike.

This led us to our second goal of presenting a conceptual framework to guide and support the choice of the most adequate tool or instrument for evaluation.

A first version of a conceptual map was presented in Sect. 5, highlighting the main constructs that characterize an EGOV evaluation instrument: context, perspective and object of analysis. The conceptual framework is oriented towards the characterization of EGOV evaluation instruments and its adequacy in a specific situation of evaluation. This framework must still be refined based on the analysis of literature from other databases such as Web of Science or Google Scholar that can complement the work done so far. The new version will also undergo a process of validation with experts through a focus group and enhancements will be continuously performed as we continue research in the area. Future work also includes a more detailed analysis of each object of evaluation regarding indicators and metrics used for its measurement. This will largely contribute for the cataloguing of the different tools and instruments available and also for the development of new ones that can complement the already existing corpus.

Acknowledgements. This paper is a result of the project "SmartEGOV: Harnessing EGOV for Smart Governance (Foundations, methods, Tools)/NORTE-01-0145-FEDER-000037", supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (EFDR).

References

1. UNECA: Measuring e-government: key policy instrument for information society development, 1–7 (2014)
2. Bikfalvi, A., De La Rosa, J.L., Keefe, T.N.: E-Government service evaluation: a multiple-item scale for assessing information quality. In: Lecture Notes Informatics (LNI), Proceedings of Series of Gesellschaft fur Informatik, vol. P-221, pp. 54–61 (2013)
3. Zhao, J.J., Zhao, S.Y.: Opportunities and threats: a security assessment of state e-government websites. Gov. Inf. Q. **27**, 49–56 (2010)
4. Miranda, F.J., Sanguino, R., Bañegil, T.M.: Quantitative assessment of European municipal web sites. Internet Res. **19**, 425–441 (2009)
5. Fitsilis, P., Anthopoulos, L., Gerogiannis, V.C.: An evaluation framework for e-government projects. In: Citizens and E-Government, pp. 69–90 (2010)
6. Ogutu, J.O., Irungu, J.K.: Citizen-centric evaluation framework for e-government systems in Kenya. The case of public service commission of Kenya Online Recruitment amp; selection system. In: 2013 IST-Africa Conference and Exhibition (IST-Africa), pp. 1–8 (2013)
7. Azab, N.A.: Assessing electronic government readiness of public organizations. Innov. Knowl. Manag. Twin Track Econ. Chall. Solut. **1–3**(8), 364–375 (2009)
8. Friedland, C., Gross, T.: Measuring the public value of e-government: methodology of a South African case study. In: IST-Africa 2010 Conference Proceedings, pp. 1–12 (2010)
9. Intergovernmental Advisory Board: High Payoff in Electronic Government: Measuring the Return on W-Government Investment (2003)
10. Bannister, F.: The curse of the benchmark: an assessment of the validity and value of e-government comparisons. Int. Rev. Adm. Sci. **73**, 171–188 (2007)

11. Kurdi, R., Taleb-Bendiab, A., Randles, M., Taylor, M.: E-Government information systems and cloud computing (readiness and analysis). In: Proceedings - 4th International Conference on Developments in eSystems Engineering, DeSE 2011, pp. 404–409 (2011)
12. Bogdanoska Jovanovska, M.: Demarcation of the field of e-government assessment. Transylvanian Rev. Adm. Sci. **2016**, 19–36 (2016)
13. Hamid, A., Sarmad, A.: Evaluation of e-health services: user's perspective criteria. Transform. Gov. People Process Policy **2**, 243–255 (2008)
14. Seng, W.M., Nishimoto, H., Nishigaki, Y., Jackson, S.: Use of satisfaction-satisfaction matrix (SSM) to evaluate Jpn. E-gov. Serv. **20**, 3–10 (2017)
15. Ostasius, E., Laukaitis, A.: Reference model for e-government monitoring, evaluation and benchmarking. Eng. Econ. **26**, 255–263 (2015)
16. Flak, L.S., Grönlund, Å.: Managing benefits in the public sector. surveying expectations and outcomes in norwegian government agencies. In: Wimmer, M.A., Scholl, H.J., Ferro, E. (eds.) EGOV 2008. LNCS, vol. 5184, pp. 98–110. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85204-9_9
17. Ünlü, N., Bayram, M.Ü., Uzun, B.: Monitoring and evaluation model of national e-government strategy and action plan: the case of turkey. In: Proceedings of 16th European Conference on e-Government (ECEG 2016), pp. 223–231 (2016)
18. Almutairi, H.: Electronic government return assessment by measuring information system usage. Electron. Gov. Int. J. **7**, 1–21 (2010)
19. Verdegem, P., Stragier, J., Verleye, G.: Measuring for knowledge: a data driven research approach for egovernment. Electron. J. Electron. Gov. **8**, 226–235 (2011)
20. Fong, S., Meng, H.S.: A web-based performance monitoring system for e-government services. In: Proceedings of the 3rd International Conference on Theory and Practice of Electronic Governance - ICEGOV 2009, p. 74 (2009)
21. Berntzen, L.: Measuring the impact of egovernment services, 54–58 (2014)
22. De Angelis, F., Polzonetti, A., Tapanelli, P.: Policy makers and performance management in e-government domain. In: Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance - ICEGOV 2010, p. 202 (2010)
23. Gupta, P.: Challenges and issues in e-government project assessment. In: Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance - ICEGOV 2007, p. 259 (2007)
24. Ziemba, E., Papaj, T., Descours, D.: Assessing the quality of e-government portals – the polish experience, vol. 2, pp. 1259–1267 (2014)
25. Shan, S., Wang, L., Wang, J., Hao, Y., Hua, F.: Research on e-Government evaluation model based on the principal component analysis. Inf. Technol., Manag (2011)
26. Awer, M.A., Esichaikul, V., Rehman, M., Anjum, M.: E-government services evaluation from citizen satisfaction perspective. Transform. Gov. People Process Policy **10**, 139–167 (2016)
27. Janssen, M.: Measuring and benchmarking the back-end of e-government: a participative self-assessment approach. In: Wimmer, Maria A., Chappelet, J.-L., Janssen, M., Scholl, Hans J. (eds.) EGOV 2010. LNCS, vol. 6228, pp. 156–167. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14799-9_14
28. Ojo, A., Janowski, T., Estevez, E.: Determining progress towards e-Government: what are the core indicators? In: Proceedings of European Conference on e-Government, ECEG, pp. 313–322 (2005)

Public Funding in Collective Innovations for Public–Private Activities

Boriana Rukanova[1(✉)], Helle Zinner Henriksen[2], Frank Heijmann[3], Siti Arna Arifah Arman[1], and Yao-Hua Tan[1]

[1] TU Delft, Jaffalaan 5, 2628 BX Delft, The Netherlands
{b.d.rukanova, y.tan}@tudelft.nl,
SitiArnaArifahArman@student.tudelft.nl
[2] Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark
hzh.digi@cbs.dk
[3] Customs Administration of the Netherlands, Laan op Zuid 45, 3072 DB Rotterdam, The Netherlands
fha.heijmann@belastingdienst.nl

Abstract. Whereas in market-driven situations the private parties have an interest in driving innovations towards implementation, in the case of public concerns, it is often the public concern that initiates the innovation process. The issue for the public funding agencies is then to stimulate idea generation and the process towards implementation and impact. However, these innovation processes are complex, as they involve a multiplicity of public and private actors with different and sometimes conflicting concerns. Thus, the benefits and business cases are not immediately clear and this makes it hard to scale beyond the proof of concept. In this paper we examine and derive lessons learned based on a longitudinal case study of a series four EU-funded projects (ITAIDE, INTEGRITY, CASSANDRA and CORE) in the international trade domain that aimed to develop digital trade infrastructure solutions (data pipelines) to address security and trade facilitation challenges. For our case analysis, we adapt and extend Bryson et al.'s framework [1] on cross-sector collaborations. We show how each of these projects covered one part of the public–private innovation trajectory, moving the innovation from the *Initial R&D* stage, to the *Showcasing* and dissemination stage to attract critical mass, towards a *Turning point* stage when the business cases for further upscaling become visible. We identify continuities (i.e. continuity of network & vision, funding and process) as well as a number of alignments as important factors that drive collective innovation processes towards implementation and impact. Further research is needed to establish to what extent these findings are applicable in other contexts.

Keywords: Innovation · Public concern · Public funding · International trade

1 Introduction

When we asked a manager from the Customs Administration of the Netherlands about the role of public funding, his response was very straightforward, namely that public funding should only be used for innovations that address public concerns and where the

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 132–143, 2018.
https://doi.org/10.1007/978-3-319-98690-6_12

private sector is not able to provide the solution due to a lack of an immediate return on investment, and that public funding should not be used for initiatives that aim to advance the interests of individual commercial companies. While it is not the goal here to argue whether this view is right or wrong, this points to one of the very important functions of public funding, that is, to help address public concerns.

It is often not easy to draw a clear line between what lies within the realm of a public concern and what does not. Public concern can be broadly seen as something that reflects the needs of society, for example safety and security, health, safe environment (including flora and fauna) and economic prosperity. These needs are sometimes equal to private commercial opportunities, but not always. And when they are not, public funding would be beneficial to drive R&D and innovation. Public funding would be valuable to enable public concerns to be directly addressed, but it can also be valuable when these public concerns are indirectly addressed, for instance when innovation leads to commercial products, which in turn generate turnover that may lead to further economic growth. In addition, a business community that does not innovate will in the long term have negative impacts on the economy, thus also on society. Added to that complexity is that in the innovation process to address some public concern, there are a multiplicity of business and government actors with sometimes conflicting goals and business models [6]. Furthermore, in the initial stage of developing innovative concepts and solutions, there are many unknowns and the parties that invest are not necessarily those that will reap the benefits. It is thus hard to take the step from proof of concept towards implementation and impact [11]. And in this process it is no longer clear where the government needs to provide support in terms of funding or other incentives, such as legislation or knowledge dissemination [4], and where to leave it to the market to do the job.

This paper is largely empirically driven. In it, we examine and derive lessons learned based on a longitudinal case study of a series of four EU-funded projects (ITAIDE[1], INTEGRITY[2], CASSANDRA[3], CORE[4]) in the international trade domain that aimed to develop digital trade infrastructure solutions (data pipelines; see [5]) to address public concerns related to security and trade facilitation. For our case analysis, we adapt and extend the framework Bryson et al. [1] on cross-sector collaborations. Based on our case analysis we show how each of these projects covered one part of the innovation trajectory moving the collective innovation from the *Initial R&D stage*, to the *S*howcasing and dissemination stage to attract a critical mass, towards a *T*urning point stage when the business cases for further upscaling become visible. We furthermore identify continuities (i.e. continuity of network and vision, continuity of funding and continuity of process) and a number of alignments as important factors that drive the collective innovation process towards implementation and impact.

The remainder of this paper is structured as follows. In Sect. 2, we present our conceptual framework, which is an adaptation and extension of Bryson et al.'s [1]

[1] See [14].

[2] http://www.integrity-supplychain.eu/.

[3] http://www.cassandra-project.eu/.

[4] http://www.coreproject.eu/.

framework of cross-sector collaboration. In Sect. 3, we present our case methodology. In Sect. 4, we present the summary of our case findings. We end the paper with a discussion and conclusions.

2 Conceptual Framework

In this paper we examine innovation processes that are intended to address a public concern and the role of public funding in such processes. A widely accepted definition of innovation is that innovation can be seen as "… an idea, practice, or object that is perceived as new by an individual or other unit of adoption" [10, p. 12]. Although this definition is a useful starting point, it refers to innovations in general and not specifically to those that are intended to address a public concern; we will come back to this point later in this section. An innovation development can be seen as a process that consists of all activities and impacts that arise from the recognition of a need or a problem, through the research, development and commercialisation of an innovation, through the diffusion and adoption of the innovation by users, to its consequences [9]. The duration of an innovation process can vary and can be as much as 15 or more years [8]. Trialability is a key concept from innovation management research [9]. Trialability means that the adoption of a technology innovation critically depends on the level at which organisations can first try out the new innovation in a confident environment and with low investments. In our cases, public funding provided such a confident environment.

More than half a century ago, [7] introduced the term collective action and suggested that such collective action is necessary among organisations to increase the speed of innovation. Looking at institutional innovation, [3] identified a number of processes related to mobilising collective action, namely framing contests, construction of networks, enactment of institutional arrangements, and collective action process. Building on the collective action view, [13] further suggest that collective innovation from initial R&D to implementation can be traced when looking at a continuum of projects. While some projects taken in isolation may appear to result in a collective action failure [2], by taking a longitudinal and cross-project perspective these projects are a necessary step in the innovation trajectory that leads these innovations from initiation to implementation [13]. In addition, the need to organise vision in the collective action for digital trade innovation is also highlighted [15]. Regarding innovations in highly regulated domains such as international trade, multiple levels also need to be taken into account [12] to reflect government influences and regulatory concerns.

[1] explain how cross-sector collaboration involving parties such as businesses, non-for-profit organisations, communities and government is becoming increasingly important for tackling complex societal challenges. Based on a thorough literature review, they develop a framework for understanding cross-sector collaborations. The framework builds on the following key concepts: initial conditions, processes, structures and governance, contingencies and constraints, and outcomes and

accountabilities.[5] The framework presents a suitable starting point for our analysis as it explicitly captures concepts that help to trace the progression from initial conditions towards outcomes and accountabilities. This is in line with our goal to illustrate the progression of public–private innovations from initiation to implementation. Furthermore, the concept of initial conditions is particularly useful, as it will allow us to capture the public concern (e.g. safety and security) and public funding that triggers collective public–private innovation processes. Bryson et al.'s [1] framework is also limited, as it does not explicitly allow an analysis of a multiplicity of projects that taken together can achieve transition towards implementation [13]. Furthermore, the multi-level nature of government is also not explicitly included [12]. The figure presented below shows our initial conceptual framework.

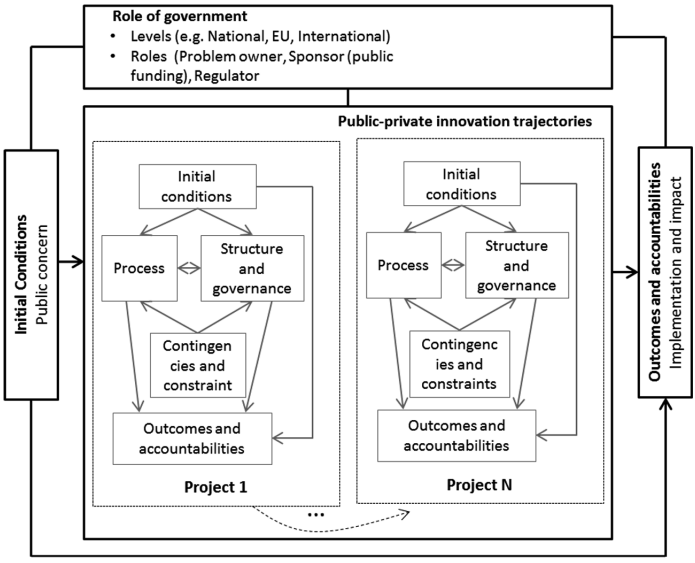

Fig. 1. Initial conceptual framework for public–private collective innovations adapting and extending the framework of Bryson et al. [1].

The above framework builds on [1] in the following way. First, inspired by [1], we look at the process from initial conditions/public concern (left-hand part of Fig. 1) towards outcomes and accountabilities (right-hand part of Fig. 1). We also incorporate the view that a succession of independent projects can together lead to the scaling up and implementation of the solutions (based on [13]). This is captured in Fig. 1 by the block labelled public–private innovation trajectories, where the trajectory contains a number of projects (1…N). At a project level, we again utilise the concepts of Bryson

[5] In Bryson et al.'s [1] framework, each of these categories is further detailed into sub-categories. For the sake of simplicity, we utilise only the high-level categories for building the adapted conceptual framework that we use in our analysis.

et al. [1] to capture how each individual project (comprising a project-specific set of actors) moves towards specific outcomes. Finally, we explicitly add the government (role and level) in our framework. The levels (national, economic zone (e.g. EU), international) are added to capture the level from which the funding or other source of government influence comes with respect to the innovation trajectories under analysis. In addition, in the conceptual framework we explicitly define three government roles to show that the agency providing the funding is not necessarily the same as the agency that owns the problem (i.e. the agency responsible for addressing certain societal problems) and drafts the regulatory and legislative frameworks.

3 Method

We conducted our case study in an interpretative, processual tradition [16]. The context of this study is a series of four EU-funded projects (ITAIDE (2005–2010), INTEGRITY (2008–2011), CASSANDRA (2011–2014), and CORE (2014–2018)) that were initiated to look for innovative digital innovations solutions to address the safety and security concern in international trade. The Customs Administration of the Netherlands played a key role in all these projects driving eCustoms innovation. A number of university and business partners also remained involved in the continuum of projects, while the network grew.

In the ITAIDE project (17 partners), the idea of reusing business data for government control purposes (also referred to as piggy-backing) was introduced and four living labs with beer, pharmaceuticals, paper and food were used to develop initial ideas and pilot these in a real-life setting [14]. While the innovations that were developed and piloted in ITAIDE showed how they solve the trade facilitation and security concerns, the business cases were not clear enough for parties to pick up the ideas further and there were also legislative constraints.

INTEGRITY (15 Partners) was started as an independent project, but through the involvement of the Customs Administration of the Netherlands some of the lessons learned from ITAIDE were also considered. In INTEGRITY, the UK Customs authority was also a partner and the idea of the data pipeline was born [5]. The data pipeline is an "an IT innovation to enable capturing data at the source" ([6], p. 14). Governments can use data pipeline information for government control purposes. In INTEGRITY, terminal operators were involved in piloting the solutions, but further upscaling and implementation of the solution was not achieved.

In CASSANDRA (26 partners), the data pipeline idea was further piloted, showcases were developed and a lot of effort was spent on disseminating the data pipeline to a wider business and government audience and on awareness building. The development of the Customs Dashboard (a special interface that the Dutch Customs authority uses to access data from the data pipeline) was initiated to allowed Customs to view data pipeline information.

CORE (81 partners) was initiated as a demonstration project with the goal to do large-scale demonstrations. Various data pipelines were piloted and business parties are now investing in developing their own pipeline solutions. One of the pipelines that is now being rolled out is a block-chain enabled global data pipeline. In addition, the

Dutch Customs authority, based on its experiences in these earlier projects, has decided to invest in its own operational Customs Dashboard that can interface with data pipelines. As such, the CORE project can be seen as a *Turning point* where business cases are becoming clear and business and government organisations are investing their own resources to move these innovations towards implementation.

The data collection took place over a period of 12 years. All the authors were involved in at least one of the projects but with different levels of involvement and roles. The authors had access to rich data. This included participation in meetings, redesign sessions, extensive interviews and document analysis in the context of the various demos, workshops and events organised by these projects. The data analysis was performed through the conceptual lens of the framework presented in Sect. 2 in a number of iterative loops. In order to deal with biases, we benefited from the fact that one of the senior researchers in the team was involved in a limited way in only the ITAIDE project and followed the other projects remotely. As such, this researcher was very instrumental in questioning the assumptions. This led to us to sharpen the analysis and the presentation of the findings.

4 Case Analysis

Figure 2 below summarises the results of our case analysis by using the conceptual framework presented in Sect. 2. In our analysis, we also identified a number of continuities and alignments that were instrumental in moving the innovation from initiation to implementation. These have therefore been added to Fig. 2 and will be explained later. It is important to highlight that unlike commercially driven ideas, where parties can come together, develop a solution and attract investors to bring an idea to the market, when it comes to innovations that are intended to address a public concern, the government has identified the concern but needs to search for ways to address it. In our case, the public concern relates to increasing safety and security, and the EU-funded projects that we discuss were intended to stimulate the development and upscaling of solutions to address safety and security challenges. Looking at the role of government in initiating the innovation trajectories under analysis in our case, this relates to the EU level (see top level of Fig. 2). In Fig. 2, we use dotted lines to indicate other levels such as national and international, as these can also be sources of public funding, but for this paper they are beyond the scope of our analysis. At the EU level, we explicitly capture that in our case while the funding was provided by DG Research, the problem ownership related to safety and security lies with other DGs (such as DG TAXUD and DG Home), which play an important role setting EU objectives related to societal goals and regulatory frameworks.

We now look at the continuum of the four EU projects and derive a number of observations by looking at the perspectives of process (P), structures and governance (S&G), initial conditions (IC), contingencies and constraints (C&C), and outcomes and accountabilities (O&A) (see Fig. 2).

Taking the process (P) perspective from the framework and looking at the continuum of projects, we see that the projects progressed through a number of stages, such as the *Initial R&D* stage in ITAIDE, where initial ideas were developed and

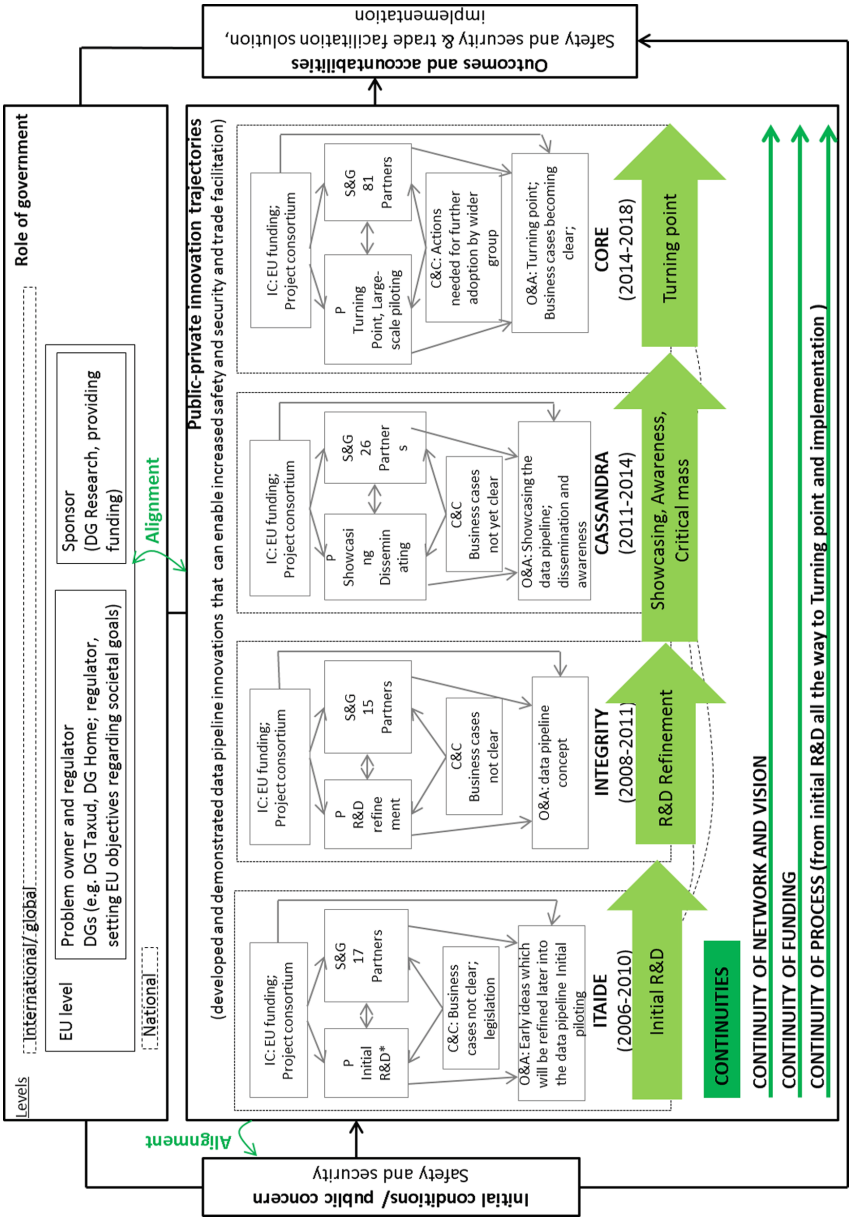

Fig. 2. Empirical model based on adapted and extended conceptual framework of Bryson et al. [1].

piloted in real-life situations; the *R&D refinement* stage, where the ideas further crystallised (the data pipeline idea in INTEGRITY) and were piloted; the *Showcasing and dissemination stage*, where showcases were developed and ideas disseminated to attract further critical mass (CASSANDRA); and the *Turning-point stage* in CORE. Here, a *'Turning point'* is the moment that individual parties take over the next step and use their own resources to further scale up parts of the solutions developed, rather than relying on public/private funding for this upscaling step.[6]

Thus, we see that in the progression towards upscaling and implementation, it was important to have *continuity of the process*[7] in terms of proceeding from one stage of the innovation process to the next stage (i.e. from the *Initial R&D* stage towards *Turning point* stage). Although in our case we do see such continuities, these are not necessarily guaranteed. If funding efforts are focused only on supporting Initial R&D and no progress is made towards the next step, this may lead to the proliferation of ideas and only limited progress towards implementation.

Looking at the structure and governance (S&G) perspective, we can trace the evolution of the network. While the network was relatively small in the *Initial R&D* stage (ITAIDE, 17 partners) and the *R&D refinement stage* (INTEGRITY, 15 Partners), the network expanded further in CASSANDRA (25 partners), where further showcases were accumulated and the idea of the data pipeline was further disseminated to create awareness. This was crucial for making the ideas tangible and getting more and more parties on board, which resulted in the further growth in the number of parties interested in committing to the idea and engaging in large-scale piloting in CORE (81 partners).

What is also key in our is case that some organisations and key people from these organisations remained involved throughout all the projects and sustained the vision. In the continuum of projects that we analysed, there were a few key individuals and organisations that had a vision and were able to carry on the ideas from one phase to another and attract critical mass and funding along the way. That was especially important in the last project, where the network grew significantly and there was a danger of losing focus. As such we see *continuity of network and vision* as key elements for the success of the innovation trajectory under analysis. Continuity of vision and network does not mean that the network remained static and was limited to the same parties; on the contrary, the network grew to include new people and organisations. This was crucial for accumulating a critical mass of parties that subscribed to the idea and were ready to co-invest in the further demos and piloting.

Looking at the initial conditions (IC) (Fig. 2), securing funding was key for each individual project. *Continuity of funding* throughout the different projects is another continuity that was essential for bringing the innovation process from initial R&D towards implementation. This was crucial, as without this continuity of funding the

[6] The R&D stage can be preceded by invention stage where some specific technology is invented and developed (e.g. the Smart container seal was developed by a commercial organisation before piloted in ITAIDE). If in the context of addressing a public concern it is discovered that there is a need to invent new technologies but there is no business drive for businesses to invest in this stage, public funding could be useful to stimulate the Invention stage as well.

[7] We do not see the process as linear, as it can evolve in a number of iteration loops.

process could have stopped at an earlier stage, before the business models were clear and the industry was ready to pick up and invest in the further implementation and upscaling of the solutions. By funding we mean both public and private funding. Had funding been stopped too early, all the earlier investments could have become sunk costs. In the continuum of projects that we discussed above, it was important that the government continued to invest, but it was even more important that in this continuum businesses became increasingly interested and co-invested millions in this process, and that the network grew indicating that more parties started to believe in the ideas. Thus, gaining commitment from businesses to participate and invest in the public–private innovation trajectory is as crucial as the funding agency that provides the public share of funding the initiative.

Regarding contingencies and constraints (C&C), the lack of clear business cases was also one of the key issues, as scaling up could not occur until some of the network partners saw clear value and were willing to further invest. The business cases started to become clear in CORE. It is still a challenge to engage more businesses and government organisations, but this process is currently being driven by pioneering businesses or pioneering governments as part of their own activities. A further challenge was how to perform data analytics on such large data volumes provided by data pipelines. This challenge still needs to be tackled.

Finally, looking at outcomes and accountabilities (O&A), the first three projects achieved project-specific outcomes. Although these outcomes in isolation were not enough to achieve the implementation of a solution to address the societal concern, they were important stepping stones towards implementation in CORE and beyond.

To summarise, we identified three types of continuities – namely *continuity of network and vision*, *continuity of funding* and *continuity of process* (from *Initial R&D* towards the *Turning point*) – as important success factors. We also identified two alignment processes that appeared important and that we discuss below.

Alignment of innovation trajectories with problem sponsor and problem owner: at the EU level, there are two DGs that are crucial for the public–private innovation trajectories: DG Research & Innovation, which provides the public funding, and the DGs that are the problem owners in the problem domain. While it is clear that there will be a lot of pressure to align the project with the DG that provides the funding (through regular review cycles), the alignment with the DGs that own the problem is also crucial for moving towards implementation in the problem domain. And we would argue that not only alignment but tight alignment with the DGs that own the problem is needed. For example, during ITAIDE, links were established with the problem owning DG and there were regular interactions, but the alignment was not very tight, as the innovation agendas had different time frames and priorities. It was therefore difficult to absorb the results of ITAIDE and move towards implementation. The innovators learned from this experience and made explicit efforts in the subsequent projects to achieve tight alignment. Ultimately in CORE a tighter alignment was achieved, which improved close communication lines and tight alignment between the innovation agenda and concerns of the problem owner DG and the efforts in CORE. However, this link is still not sufficient to align the policy agenda with the outcome of research. That alignment is crucial, especially in highly-regulated domains, such as the area of Customs. The support of the regulator is crucial for ensuring the adoption and upscaling of the

innovations and realising further impact. Without such support, even the best solutions on paper will be hard to implement.

Alignment with the public concern and with the innovation trajectory and vision: we found out that it was necessary to constantly match the innovation process steps covered in each project to the source (the reason, the initial conditions/public concern) of funding. We found out, in particular in CORE, due to the many new actors that entered the project it was initially difficult to keep them focused on the initial concept, which made it difficult to keep track with the initial idea. Thus, continuous checks and balances are crucial, especially as the network grows and more actors are involved in piloting. Related to alignments, it is our view that in all these steps it is possible to find a balance between public and private interest, shared benefits: trade facilitation and safety/security are two sides of the same coin. This, of course, will not always be the case in R&D related to public concerns, but in the case of the data pipeline concept this was a fruitful outcome.

5 Discussion and Conclusion

In market-driven situations, innovations can start with inventions where some party develops a new technology and then looks for investors who have an interest in this innovation and will enable it to grow. In the case of public concerns, it is not the private market but the public concern that initiates the innovation process. The task of public funding agencies is to stimulate the generation of ideas and then stimulate the process of moving these ideas towards implementation and impact.

In this paper, we examined and derived lessons learned based on a longitudinal case study of a series four EU-funded projects (ITAIDE, INTEGRITY, CASSANDRA, CORE) in the international trade domain that aimed to develop digital trade infrastructure solutions (data pipelines) to address safety and security challenges. In the approach followed in these projects, trade facilitation was included as an additional public concern that was taken into account in shaping the solutions. For our case analysis, we adapted and extended Bryson et al.'s [1] framework on cross-sector collaboration. We extended this framework by explicitly acknowledging that the progression of innovations from Initial R&D towards implementation can span multiple independent projects that individually can achieve limited outcomes but together can be seen as part of innovation trajectories that move innovations towards implementation. Based on our case analysis, we identified continuities (i.e. continuity of network and vision, continuity of funding and continuity of process (from Initial R&D towards Turning point and implementation)), as well as two alignments of the innovation trajectory (i.e. alignment with the project owner and project sponsor, and with the public concern), as important success factors.

As discussed, trialability is a key concept from innovation management research [9]. Trialability means that the adoption of technology innovation critically depends on the level at which organisations can first try out new innovations in a confident environment and with low investments. In our cases, the public funding provided such a confident environment, where a complex network of public and private actors could experiment, develop solutions, test them in real life and learn. In this process, the

uncertainties and ambiguities were gradually reduced, and the showcases attracted critical mass and further funding. This process made it easier to reach a Turning point where the business cases gradually become more clear, which made it easier for parties to take the next step and proceed with further implementation and upscaling with own investments.

Although in the projects that we analysed we saw continuity of network and vision, continuity of funding and continuity of process, such continuities are not guaranteed. These continuities need to be carefully managed as they may be endangered when moving from one project to another. As a result, the vision, as well as the earlier efforts and investments, may be lost or insufficiently utilised. For example, if the funding and efforts are mainly focused on the R&D stage – or are devoted to more fundamental research or broadened to adjacent fields – and no steps are taken towards the Showcasing and dissemination stage to build awareness, or towards the Turning point stage, then it is likely that a lot of ideas will be generated but limited impact in terms of scaling up and adoption will be achieved. Similarly, if there is continuity of funding but not of network and vision, this may lead to a situation in which know-how and expertise are not efficiently utilised when moving towards implementation.

In discussions it was suggested that after the Initial R&D stage, the government could step forward and assume the role of an investor by taking the risk and investing, but if the project is successful the government would be one of the parties realising the gains and with the gains, the financing of a next project can be assured. Whereas different scenarios may be possible, we consider that continuity of vision and network, continuity of funding and continuity of process from Initial R&D to Turning point, as well as the alignment processes that we identified, are crucial for moving towards implementation and impact.

This study was largely empirically based and the lessons learned are limited to the four projects under analysis. Further research is needed to establish the applicability of the findings in other context and domains. However, the lessons learned can be insightful also for other initiatives where public funding is used to facilitate the development and upscaling of innovative solutions to address some public concern.

Acknowledgement. This research was partially funded by the CORE Project (nr. 603993), which is funded by the FP7 Framework Program of the European Commission. Ideas and opinions expressed by the authors do not necessarily represent those of all partners.

References

1. Bryson, J.M., Crosby, B.C., Stone, M.M.: The design and implementation of cross-sector collaborations: propositions from the literature. Public Adm. Rev. **66**(1), 44–55 (2006)
2. De Reuver, M., Verschuur, E., Nikayin, F., Cerpa, N., Bouwman, H.: Collective action for mobile payment platforms: a case study on collaboration issues between banks and telecom operators. Electron. Commer. Res. Appl. **14**(5), 331–344 (2015)
3. Hargrave, T.J., van de Ven, A.H.: A collective action model of institutional innovation. Acad. Manag. Rev. **31**(4), 864–888 (2006)

4. Henriksen, H.Z., Andersen, K.V.: Diffusion of e-commerce in Denmark: an analysis of institutional intervention. Knowl. Technol. Policy **17**(2), 63–81 (2004)
5. Hesketh, D.: Weaknesses in the supply chain: who packed the box. World Customs J. **4**(2), 3–20 (2010)
6. Klievink, J., et al.: Enhancing visibility in international supply chains: the data pipeline concept. Int. J. Electron. Gov. Res. **8**(4), 14–33 (2012)
7. Olson, M.: The Logic of Collective Action. Harvard University Press, Cambridge (1965)
8. Ortt, J.R.: Understanding the pre-diffusion phases. In: Gaining Momentum: Managing the Diffusion of Innovation, pp. 47–80 (2010)
9. Rogers, E.M.: Diffusion of Innovations. Free Press, New York (1995)
10. Rogers, E.M.: Diffusion of Innovations, 5th edn. New Press, New York (2003)
11. Rukanova, B., Henningsson, S., Henriksen, H.Z., Tan, Y.H.: Digital trade infrastructures: a framework for analysis. Complex Syst. Inform. Model. Quart. **14**, 1–21 (2018)
12. Rukanova, B., van Stijn, E., Henriksen, H.Z., Baida, Z., Tan, Y.H.: Understanding the influence of multiple levels of governments on the development of inter-organizational systems. Eur. J. Inf. Syst. **18**(5), 387–408 (2009)
13. Rukanova, B., de Reuver, M., Henningsson, S., Tan, Y.H.: Overcoming blockages in collective innovation in digital infrastructures. In: Proceedings of ECIS 2017 (2017). http://aisel.aisnet.org/ecis2017_rp/71
14. Tan, Y.H., Bjorn-Andersen, N., Klein, S., Rukanova, B.: Accelerating Global Supply Chains with IT-Innovation. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-15669-4
15. van Stijn, E., Rukanova, B., Wensley, A., Tan, Y.H.: Moving an eInnovation from a Living Lab to the real world: politically savvy framing in ITAIDE's Beer Living Lab. In: Proceedings of the 22nd Bled eConference, Bled (2009)
16. Walsham, G.: Interpreting Information Systems in Organizations. Wiley, Chichester (1993)

Ontology Based Data Management

A Study in a Brazilian Federal Agency

Márcia Myuki Takenaka Fujimoto[1(✉)] and Edna Dias Canedo[2]

[1] Ministry of Transparency and Comptroller General, Brasília, Brazil
marciamyuki@gmail.com
[2] Computer Science Department, Professional Masters in Applied Computing, University of Brasília - UnB, Brasília, Brazil
ednacanedo@unb.br

Abstract. The Ministry of Transparency and Comptroller General is the agency of the Federal Government in charge of assisting the President regarding the treasury and public assets and the government's transparency policies. The Agency takes care of active transparency mechanisms on federal public resources, that is, improves actions related to information the State must disclose, without being demanded for. It establishes ways of assuring information will be appropriated and effectively used by society, including through web applications for citizens. Thus, the Agency is part of a governmental environment in which the complexity of data management involves improvement of information quality and interoperability between systems, as a consequence the need of a capable model to manage all of its data assets rises up. It's necessary to organize and implement a data management capability that allows understanding (semantic), finding and sharing data. This article describes a research study directed to data management under ontological approach, which proposes an Enterprise Information Architecture Model in the form of a conceptual layer-based data prototype, taking into account both academic and industry-driven studies.

Keywords: Information and Communications Technology (ICT)
Data management · Data integration · Ontology · Public company

1 Introduction

The Ministry of Transparency and Comptroller General (CGU) is the agency of the Federal Government in charge of assisting the President of the Republic regarding the treasury and public assets and the government's transparency policies. These tasks are out by way of public audits, fraud deterrence procedures, and other sort of internal control, corruption prevention and ombudsman activities. The performance of its various activities is often done through the consolidation of information, cross-referencing data, data mining and open data availability. Such activities require a holistic and integrated data assets view. The existing data management structure is elementary and has not shown to be effective in meeting CGU requirements, resulting in several challenges.

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 144–154, 2018.
https://doi.org/10.1007/978-3-319-98690-6_13

The main input of the many tasks of the CGU is the data from three main sources: self-owned systems to support their final activities, databases of structuring systems of the Federal Government and external data provided by systems of institutions connected to the Government. The data unchecked proliferation in the Agency leads to problems in data governance and management. Some of that is due to the system's architecture being organized in "silos", meaning it's made up of a myriad of data sources, independent and distributed, each serving a specific application [1]. Thus, even though there is a significant amount and variety of data sources, there are basic questions which are hard to answer, such as:

Q1. How many and which are the data sources that contain information regarding the organizational structure of government, regardless of rank or power?
Q2. What are the sources from the CGU systems that contain citizens' data and can be integrated?
Q3. What are the data sources to be searched in order to obtain information about all CGU control actions in a particular Agency?

There are challenges in finding, integrating and improving organizational data quality. As a consequence, interoperability among systems is jeopardized. Any solutions to get over such difficulties need management capabilities focused on semantic integration, both intra- and inter-organizationally. Such solutions demand data sources labeling with business and technical metadata, logically ordered for these. Aware of this situation, CGU launched in 2014 a working group to study the Cobit 4.1 PO2 Process - Define the Information Architecture [10] for a coming implementation. Considering that the implementation has not occurred so far, there is an opportunity to a new one PO2 process implementation proposal. Cobit 4.1 presents two control objectives directly related to the data understanding and sharing: PO2.1 - Enterprise Information Architecture Model; PO2.2 - Enterprise Data Dictionary and Data Syntax Rules.

The data integration and sharing issue is core to the PO2 process. Because of this, this study considered holistic approaches related to ontology based data management - OBDM to ensure the true data integration. This study intends to, more specifically, present an information architecture proposal, in the form of a conceptual ontology based data model, supported by a data dictionary (DD) solution architecture coherent with the recommended approach. The information architecture will generate a prototype for evaluation purposes.

The remainder of the paper is organized as follows. Section 2 presents the background in which the theoretical bases for conducting this research is raised. Section 3 summarizes the research and architecture development methodologies. Section 4 presents the Case Study and the preliminary results of this work. We conclude the paper presenting some final remarks in Sect. 5.

2 Background

The systematic review carried out, described below, resulted in the choice of the Ontology based data management - OBDM proposal [2] and the holistic multi-domain architectural structure by [3]. They have a great similarity of purpose with this work

with regard to an overview of the enterprise as a whole and to the data integration and sharing. Other sources for the study include some Federal Government initiatives related to the data management field.

2.1 Ontology Based Data Management

The OBDM approach can be seen as a way of integrating information in which the global schema of data is substituted by a conceptual model from the domain of interest to a given organization formally specified in an ontology. The architecture to which the main idea of OBDM relates is divided in three levels: ontology, data sources and mapping between these two. More specifically for OBDM, ontology is the formal description of a domain of interest to a given organization, expressed through its relevant concepts, such as concepts' attributes and the logical affirmations that characterize knowledge on the domain. The data sources, in turn, are repositories accessible by the organization where the data domain is stored. Frequently, these repositories are numerous and heterogeneous, with each one being managed and kept independently from the others. The level of mapping is a precise specification of the correspondence between data kept in the sources and elements of the ontology [2].

This division in layers has three main advantages. Firstly, the ontological layer which, by making the representation of the domain explicit, allows the re-usability of the acquired knowledge and a unified description of underlying data sources. Secondly, the mapping layer explicitly specifies the relationships between the concepts of domain (ontology) and the data sources. The ontology and mapping corresponding to the data sources provide a common element for the documentation of all the data in the organization, with obvious advantages for the governance and the management of the information systems. The third advantage relates to the extensibility of the system, which doesn't require to fully integrate the data sources at once. Instead, after building a skeleton of the domain model, new sources or elements therein can be incrementally added [2].

The study presented by [13] regarding semantic databases details how the mapping between sources and ontology occurs. The main idea is to map tables and attributes of a Database (DB) for a determined ontology. This ontology must be formal in regards to the implementation of a transitive hierarchy "is-a", which connects all concepts.

2.2 Multi-domain Reference Architecture

Fitzpatrick [3] explains that the ontology must distinguish between domain knowledge, that may be extra organizational, versus localized application level knowledge. Besides, he explains the idea of the criterion of orthogonality that is applied to the structure proposed. The criterion of orthogonality is defined as the requirement of basing a newly created ontology on one or more existing ontologies. This practice, if generalized, would help reduce the "silo" effect in the development of ontologies [19].

Fitzpatrick [3] proposed reference architecture ontology structure is composed of the top-to-bottom ontologies (see Fig. 1). In light of the criterion of orthogonality, the proposed multi-domain data integration ontology subsumes in respect with the foundational ontologies. Domain specific ontologies are subsumed to the multi-domain

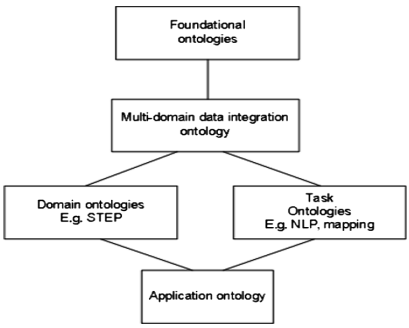

Fig. 1. Reference architecture ontology structure

ontology proposed. Then, the ontology structure comprises generic task ontologies and the structure is completed with application ontology to support domain specific tasks.

The structure proposed by [3] fits into the ontology layer of the OBDM architecture and allows the implementation of master data management (MDM). The multi-domain data integration layer comprises master data ontologies, that is basically fundamental data for all business transactions, essential cross-enterprise assets that contribute to many management paradigms [3]. This reference architecture by allowing MDM implementation ensures a collaborative environment between the many business paradigms or processes.

2.3 Federal Public Administration Initiatives

The Framework of Enterprise Architecture for Interoperability in Governance Support (FACIN) seeks to support the Brazilian Digital Governance Strategy. Through the establishment of the Enterprise Architecture and of interoperability standards, the FACIN will act as a reference for the many agencies of the Federal Public Administration FPA[1] [6].

The FACIN is made up of nine visions (domains), in which are the Data vision and the Applications vision. Structurally, it's composed of four main parts, among them:

- Content Model (CM) - Describes the structure of related elements that describe generic models for representation of Federal Public Administration organizations. CM is specified through a conceptual model presented by a diagram that joins metadata according to two perspectives [4]: Data Description and Data Sharing. The FACIN conceptual Data Description Model is supported by the Entity Relationship (ER) modelling and it encompasses semantic, syntactic and technical metadata.
- Reference Model (RM) - Describes standards, guides and best practices for the development of the FACIN from the strategic to the operational levels, focusing in the integration and construction of the Government vision as a whole [6]. RM

[1] FPA corresponds to the set of agencies of the direct, autarchic and foundational administration.

points to the observation and care with the de definition of master data as one of the critical factors of success in implementation of the data vision. Besides that, it stipulates the utilization of the Global Data Model (GDM) and the Controlled Vocabulary of Electronic Government (CVEG) as standards.

Global Data Model - GDM is a model created with the idea of enabling and ensuring the integrability of information generated for the decision process. Essential data and process modeling is used to create GDM model. It is called essential because it considers only the relevant information to the understanding of the business domain, discarding operational or technological details. Essential modeling acts similarly to a reverse engineering process of description of the database of the legacy systems [18]. The GDM aims to schematically represent the data treated by the information systems, seeking to create an integrated metadata platform. Therefore, the GDM produces some artifacts related to the data semantic and process understanding, among them, a Complete Integration Diagram. It's a vision that puts all ER entities together, a global schema of data that is a big model in ER style and highlights the common entities between information systems.

Vocabularies and Ontologies of the Electronic Government (e-Vog) intend to be a set of standards, tools and methodologies to allow: information exchange under semantic agreement, so as to favor data matching from various sources; elicitation of the tacit knowledge from government business areas by using ontology as a tool for conceptual modeling [17]. The e-Vog is a initiative under construction and has not yet made its products available, except the Repository of Vocabularies and Ontologies of Electronic Government, a site to access all ontology references of the Electronic Federal Government [17]. As stated before, FACIN indicates GDM, an ER model, to be adopted in the implementation of the data vision in the federal agencies, seeking to make feasible an integrated metadata platform. Besides that, FACIN points to take care of the de definition of master data. On the other hand, the e-Vog initiative indicates the use of ontologies for conceptual data modeling as being in this aspect more aligned to the proposal of this work.

These FPA initiatives are still evolving, but currently they do not provide ready and appropriate models for CGU needs. Either way, they do offer some directions that can be observed for the development of CGU information architecture and DD solution.

3 Methodology

One of the first activities of this study was a systematic literature review about studies that dealt with ontology for data management and, after that, the analysis of some that were selected. The systematic review process was based on the Kitchenham e Charters' guide [12]. The main results of the systematic literature review allow us to conclude that:

- the OBDM has received increasing attention since 2006, with numerous ontological data model proposals with an adjacent technical architecture applied to real problems;

- the reporting of implementation experiences corroborates the proposed models applicability;
- the majority of the studies with a model proposal deal with a specific domain;
- most of the authors indicate the need for expansion of use and evolutions in the presented models.

Considering that the ICT field is somewhat lacking in fundamental theory, a mixed research approach is recommended, with a theoretical base and practical verifications, especially in regards to data integration [7]. Knowledge can be developed (drawn) from academic research and (also) practice [7]. Based on that, other sources for the study include guides for best Information and Communications Technology (ICT) governance practices as well as some related Federal Government initiatives. The ICT guides considered are: Control Objectives for Information and Related Technologies - COBIT 4 [10] and the Togaf 9 [9, 20].

As regards the development of the information architecture, the ontology based conceptual data model, it followed the informal phase of Enterprise Approach of ontology construction [21]. For this, it was decided to use the questions and problems mentioned in the introduction 1 of this work as competency questions to identify the ontology based model scope, a list of potentially relevant concepts [11]. One of the ways to determine the scope of the ontology is to sketch a list of questions that a knowledge base based on the ontology should be able to answer, competency questions [8].

As regards the development of the DD solution architecture, it followed customized steps of Togaf Architecture Development Method - ADM - phase "A. Architecture Vision" [9]. The TOGAF ADM is a generic method that describes a method for developing an enterprise architecture. ADM is phases cycle composed, ranging from the phase "A", of the initial architecture view, to the phase "H" which identifies further necessary changes. The phase "A" objective is to create a vision of the architecture proposal (version 0.1). It intends to help the approval decision and, also, to help the understanding of its impacts.

4 Case Study and Preliminary Results

The systems and activities that assist CGU functions deal with varied and complex data. The proper use of CGU data sources requires effective management by a common element for the documentation of all the data in the organization: the ontology-based data model.

4.1 Ontology-Based Data Model

The current stage of this research did not allow the development of a proper CGU ontology, in which the concepts of its elements are shared by the users of the proposed system. Therefore, we have currently opted for the proposal of an ontology based conceptual data model, which can evolve to become an ontology on a future stage. The development of the conceptual data model and its prototype for this research was done through a few options: i. the adoption of OBDM approach of integrating information;

ii. the customization of the multi-domain model [3], carefully observing the FACIN recommendation of taking care of the sovereignty of information, master data [5]; iii. the use of CGU domain based concepts, CVEG based concepts, in compliance with the MR-FACIN recommendation [5], and other available ontologies; iv. the use of the Protégé [16] tool, web version, to draft the ontology based data model prototype.

The basic structure of CGU proposed model comprises a multi-domain data integration layer - MD, a business domain layer - BD, a generic task layer - TA and an application layer - AP. In the proposed model, business data domains are subsumed in respect with MD layer data domains. The application data domains subsume BD layer data domains and TA layer data domains. The MD layer data domains were selected from some enterprise ontologies used as references for this work, even they not belong to governmental area. This is possible due to the multi-domain concepts generality.

Figures 2, 3 and 4 were taken from the prototype structure of the ontology-based conceptual model built on the WebProtégé tool. Figure 2 identifies data domains that will compose the MD layer. They address the fundamental concepts, some already known in data modeling, that allow the systems interoperability. It means that this layer can help answer questions about master data like Q1 and Q2 in the Sect. 1.

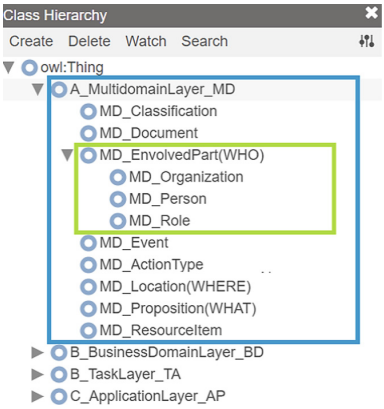

Fig. 2. Data domains for the multi-domain data integration layer

Figure 3 left side identifies some concepts that make up BD layer, they represent functions performed by the agency and its specializations, which have been derived from the information in [14]. This layer is consistent with the GDM view as it addresses only the scope of the business domain. Figure 3 right side shows how the data domains are related to describe the CGU functions in a general way: Subject (involved parties) + action (type of action from the agency) + object (resource item or other) and complement (locale and others). All the elements which are necessary to the description in this general way must be found in the MD layer or in the in the BD Layer. This layer can help answer questions like Q3 in the Sect. 1.

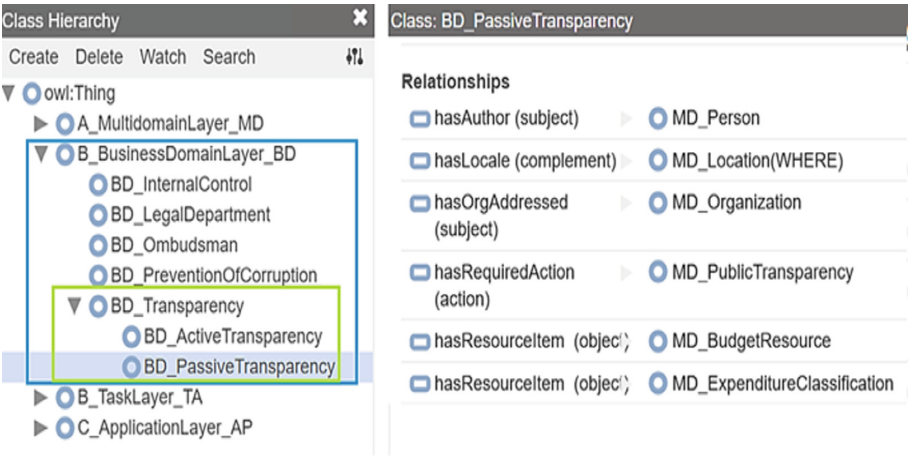

Fig. 3. Business domain layer and concepts relationships

Figure 4 left side identifies some elements that make up the TA layer. The layer is organized by subjects related to tasks that are independent from the business domain, that could become reusable corporate services. The listed subjects are simply illustrative, able to evolve as applications are cataloged. The TA data domains may subsume from MD layer data domains, for example "Document", or may be a proper subject like "File". This layer can support the reuse of general services and data sharing, for instance, user's access authorizations task in "Security" subject.

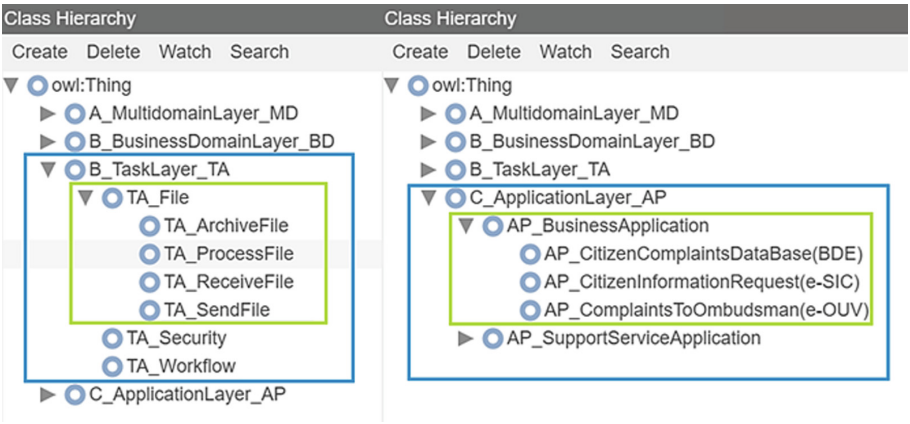

Fig. 4. Task layer and application layer

Finally, the AP layer will have two divisions: BusinessApplication and SupportServiceApplication. The CGU applications that most meet the business functions will be under the first division and the ones destined to offer support services will be

under the last one. Figure 4 right side presents the two main divisions of the AP layer and some examples of existing applications in the Agency. The data domains of each application are subsumed in respect with BD domains or TA domains.

The proposed conceptual model, top-down, is made of layers. Its is different from GDM type models that are created by ER models composition of applications that interact between themselves, with a bottom-up approach. Nevertheless, each integration point identified in the GDM type models must find a corresponding concept or relationship in the proposed model MD layer. Besides that, each data source of the CGU applications, when submitted to reverse engineering, should find a corresponding concept or relationship in any layer of the proposed model, as a way to verify its completeness.

Over the rest of the research project, the prototype will be evolved and data sources from two CGU applications will be mapped to the proposed model concepts. The directions presented in [13] regarding semantic database will be used to map the applications database tables to the proposed ontology based model. Next, the prototype will be explored by CGU experts (like database administrators, application developers, business intelligence analysts and business specialists) to answer a qualitative survey in order to validate the proposed model.

4.2 Architecture for Data Dictionary Solution

The architecture considered suitable for CGU needs is presented in the Fig. 5 and its main elements are described in the Table 1. The proposal is oriented by OBDM's base idea [2] of separation between ontology and data source, with their respective mappings. This version 0.1 architecture vision, that covers business, application, data and technology domains, was draft in an Archimate tool. Archimate is an open enterprise architecture modeling language to support the description, analysis and visualization of architecture within and across business domains [15]. In future experiment stages in

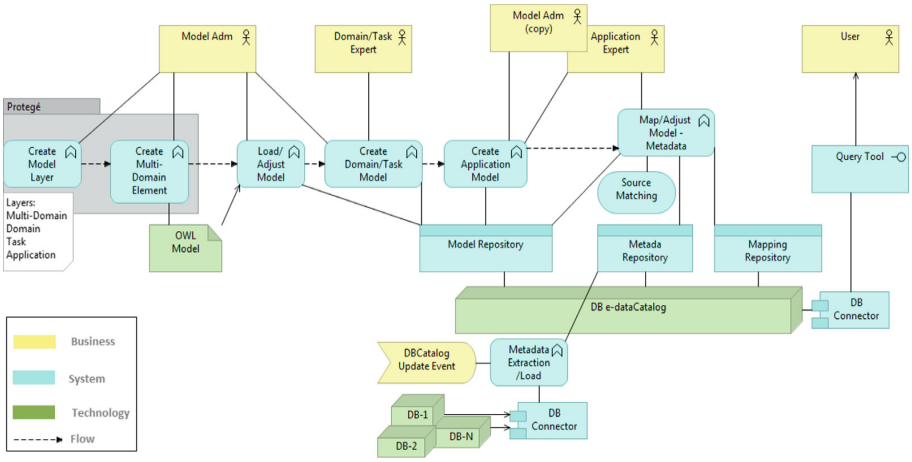

Fig. 5. Proposed architecture for DD solution

this project, there might be a simplification in some components, such as the removal of the extraction and automatic load of WebProtégé [22] ontology elements to the ontologies repository.

Table 1. Architecture components

Element	Description
Model Adm	Actor responsible for maintaining the model basic structure and assists other managers so as to keep it coherent
Domain/task expert	Specialist actors that describes a domain or task/service
User	Actor who uses the DD information
Create model layer	Model structure creation function
Create MD element	Master data domains creation function
Load/adjust model	Model data extraction and relational repository loading functions. Adjustments are allowed to the Model's Administrator
Create domain/task model	BD and TA data domains creation function. It allows concepts and relationships creation
Create application model	AP data domains creation function. It allows concepts and relationships creation
Metadata extraction/load	Function that automatically extracts technical metadata from structured data sources to load on the metadata repository
Map/adjust model - metadata	Function that allows the mapping between structured data sources (tables) and the model concepts. It can uses table matching solutions, through structure or content analysis

5 Conclusion

The proliferation, sometimes uncontrolled, of electronic data in the government area, as well as in other areas of society, is a fact. As control mechanisms and the transparency of government actions evolve, becoming more complex, so does the need of a robust management capacity of this fundamental asset. The recent tendency of data management with ontology support provides a promising path to solve problems in data understanding, localization, reuse and integration, improving the interoperability between systems and the quality of the information created.

This study considers the academic advances and industry initiatives in order to elaborate a proposal for the implementation of the management of multi-domain data with an ontological approach for the CGU, using quantitative and qualitative research. We hope that the proposal can help the CGU PO2 process – Define the Information Architecture - effective implementation, as well as aligning the CGU initiative to the ones existing in the FPA.

The prototype of the conceptual model and the proposed architecture should be a little more detailed and validated with its application in a reduced scope of CGU applications, to be defined at a later date. Besides, there is the concern of doing a more minimalistic detailing with the maintenance of a central structure that can be adapted to other FPA agencies.

References

1. Daraio, C., et al.: Sapientia: the ontology of multi-dimensional research assessment. In: ISSI (2015)
2. Daraio, C., et al.: Data integration for research and innovation policy: an ontology-based data management approach. Scientometrics **106**(2), 857–871 (2016)
3. Fitzpatrick, D., Coallier, F., Ratté, S.: A holistic approach for the architecture and design of an ontology-based data integration capability in product master data management. In: Rivest, L., Bouras, A., Louhichi, B. (eds.) PLM 2012. IFIPAICT, pp. 559–568. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35758-9_50
4. Governança Corporativa, G. T.: FACIN - Framework de arquitetura corporativa para interoperabilidade no apoio à governança - modelo de conteúdo v1.5 (2017). http://www.participa.br/articles/public/0056/4482/ModelodeConteudov1.5.pdf. Accessed 01 Jun 2018
5. Governança Corporativa, G. T.: FACIN - Framework de arquitetura corporativa para interoperabilidade no apoio à governança - modelo de referência v3.1 (2017). http://www.participa.br/articles/public/0056/4483/Modelo_de_Referencia_v3.1.pdf. Accessed 01 Jun 2018
6. Governança Corporativa, G. T.: FACIN - Framework de arquitetura corporativa para interoperabilidade no apoio à governança - visão executiva v2.1 (2017). http://www.participa.br/articles/public/0056/4481/FACINVisaoExecutivav2.1.pdf. Accessed 01 Jun 2018
7. Gregor, S.: Building theory in the sciences of the artificial. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, p. 4. ACM (2009)
8. Grüninger, M., Fox, M.S.: Methodology for the design and evaluation of ontologies. Citeseer (1995)
9. Haren, V.: Togaf version 9.1. Van Haren Publishing, Zaltbommel (2011)
10. ISACA, C.: Cobit R 4.1 process assessment model. ISACA IL, USA (2011)
11. Jones, D., Bench-Capon, T., Visser, P.: Methodologies for ontology development (1998)
12. Kitchenham, B., et al.: Systematic literature reviews in software engineering - a tertiary study. Inf. Softw. Technol. **52**(8), 792–805 (2010)
13. Köhler, J., Philippi, S., Lange, M.: SEMEDA: ontology based semantic integration of biological databases. Bioinformatics **19**(18), 2420–2427 (2003)
14. Ministério da Transparência e Controladoria-Geral da União – competências page. http://www.cgu.gov.br/sobre/institucional/competencias-e-organograma. Accessed 01 Jun 2018
15. Open Group Archimate Overview. http://www.opengroup.org/subjectareas/enterprise/archimate-overview. Accessed 01 Jun 2018
16. Protégé. https://protege.stanford.edu/. Accessed 20 May 2018
17. Repositório de Vocabulários e Ontologias de Governo Eletrônico. http://vocab.e.gov.br/. Accessed 01 Jun 2018
18. SERPRO: Guia metodológico para integração de dados e processos modelo global de dados (2017). http://modeloglobaldados.serpro.gov.br/modelo-global-de-dados/guia-metodologico. Accessed 20 May 2018
19. Smith, B.: Ontology (science) (2008)
20. Togaf Version 9.2. http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html. Accessed 01 Jun 2018
21. Uschold, M., King, M.: Towards a methodology for building ontologies (1995)
22. WebProtégé Tool page. https://webprotege.stanford.edu/. Accessed 20 May 2018

Towards the Implementation of the EU-Wide "Once-Only Principle": Perceptions of Citizens in the DACH-Region

Cigdem Akkaya(✉) and Helmut Krcmar

Chair for Information Systems, Technical University of Munich, Munich, Germany
{cigdem.akkaya, krcmar}@in.tum.de

Abstract. This paper presents selected findings from a recent empirical research conducted in the context of modernization of public administrations. E-government adoption has reached a stagnation point over the last few years in the DACH region. The European Commission has been working intensively on projects aiming to overcome the existing digital barriers between EU Member States. Citizens repeatedly provide the same personal information to different public authorities, which causes frustration and inefficiency. The Once-Only Principle suggests that citizens should have the right of providing information to public authorities only once and that the provided data will be exchanged between national authorities of the EU Member States. By signing the Tallinn Declaration in October 2017, EU Member States have already committed to implement this principle. Sharing personal data of individuals between public authorities within national boundaries as well as with the public administrations of other EU Member States would assuredly ease-up data provision and increase efficiency. Yet, higher convenience comes at a cost of data protection and privacy, which becomes highly critical when sensitive personal data is involved. From this standpoint, a particular emphasis needs to be placed on understanding expectations, sensitivities and privacy related concerns of citizens, which is argued to be one of the key drivers behind the adoption of G2C e-government initiatives.

Keywords: E-government · Single Digital Gateway · Once-Only Principle
Privacy and data protection · DACH region

1 Introduction

Utilization of e-government by transforming public service delivery is quite promising in terms of eliminating long queues, improving efficiency and providing higher convenience to citizens. Improving public service delivery by application of digital technologies is high on the political agenda of many EU countries. Despite considerable investments, success of most Government-to-Citizen (G2C) e-government initiatives remain far below expectations. It is not realistic to expect a sudden take-off without analyzing the underlying barriers, which hold people back from an intensiver usage of digital public services. Although the sensitivities of nations may differ, privacy and

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 155–166, 2018.
https://doi.org/10.1007/978-3-319-98690-6_14

data protection concerns are one of the widely recognized barriers to adoption of services in online contexts [1, 2]. For example, lack of confidential handling of data, fear of becoming a "transparent citizen", fear of data theft and lack of information about the usage of personal data shape the privacy related concerns and reluctance towards using e-government services in Germany [3, 4].

The issue of privacy becomes even more important considering the new initiatives of the European Commission (EC) that aim to modernize digital public services and improve their cross-border availability. The motivation of the project is clear. Increasingly more people cross national borders to live, study, work or retire. This creates additional documentation and paperwork. For instance, if a person works in one EU country but lives in another one (a.k.a 'cross-border commuters'), his or her (further "he") social security contributions are likely to be covered by the EU country where he works but if he loses his job, he may need to apply for benefits in the country where he lives. The necessary paperwork may be less in his original country. However, public authorities in the second country do not likely have any access to his personal records or any other previous information. Providing all this information to government agencies of another country is not only overwhelming and time consuming for individuals, but results also in additional financial burdens such as translation and notary costs.

In order to eliminate these problems and enhance cross-border activities within the EU, the EC has introduced a concept called the "Once-Only Principle" (OOP). With this concept, the provision of the same personal information to different government offices would be theoretically eliminated. Its implementation is quite essential due to its importance in successful cross-border digitalization. Indeed "enabling mobility of citizens and businesses[1] by cross-border interoperability" has been stated as one of the three policy priorities of the eGovernment Action Plan 2016–2020 [5]. Besides making life of citizens easier, exchange of information between government offices of different EU Countries could support fighting terrorism and crime.

In October 2017, by signing the Tallinn Declaration on eGovernment, the ministers in charge of eGovernment policy and coordination from 32 countries of the EU and the EFTA have committed to "take steps to identify redundant administrative burden in public services and introduce once-only options for citizens and businesses in digital public services by collaboration and data exchange across our administrations at national, regional and local level as well as with other countries for cross-border digital public services" ([6], p. 4). The declaration acknowledges the protection of personal data and privacy by introducing the General Data Protection Regulation (GDPR). However, this regulation is focused mainly on interactions between companies and individuals rather than citizens and governments.

Data sharing and re-use of data, however, bring important questions about data protection. In some countries, data exchange between national public authorities is perceived with major resistance due to strong data protection regulations [7], therefore, individuals would be likely to be more reserved towards data exchange between public

[1] The OOP is introduced for both citizens and businesses but we focus on citizens as G2B e-government goes beyond the scope of this paper.

authorities of all EU Member States. The EU level documents are kept mostly quite high-level without providing much detail about implementation and their consequences for end-users. Neither the Tallinn Declaration [6] nor the eGovernment Action Plan 2016–2020 [5] specify the type, details or scope of information to be shared within the EU Member States. In fact, they mostly focus on benefits of this approach and potential savings to be achieved through its utilization.

We argue that design and implementation of such initiatives needs to be done with utmost care by taking the concerns and sensitivities of the individuals into account. To the best of our best knowledge, academic research on G2C e-government has not yet tackled the citizen perspective of the OOP with empirical analyses. Other than a few conceptual papers [8–12], a study focused on its legal perspective [13] and project reports [14][2], no empirical study reflecting the citizens' perspectives on OOP could be found. As the implementation of the OOP is still in its infancy, we aim to gain early insights, which could possibly be taken into consideration when planning its implementation. In particular, we focus on understanding privacy related concerns of the DACH region citizens. The following questions guide this research:

RQ1. How has the rate of G2C e-government adoption annually changed between the years of 2014 and 2017 in the DACH region?
RQ2. What are the characteristics of a modern government agency from the perspective of citizens in the DACH region? Do they consider the OOP as a characteristic of a modern government agency?
RQ3. What is the opinion of citizens in the DACH region regarding sharing data between public authorities?
RQ3.1. What is their opinion regarding data sharing between public authorities within their own country?
RQ3.2. What is their opinion regarding data sharing with public authorities of other EU Member States?

As we employ nationwide representative samples, the results are generalizable to the whole population.

2 Background and Literature Review

2.1 Digital Single Market Strategy and Digital Single Gateway

To facilitate the operation of the EU Single Digital Market, the EC has introduced the Single Digital Gateway Strategy in May 2015 [15]. This strategy aims to unlock the full potential of the European Single Market. In particular, it foresees the free movement of persons, services and capital within the EU, irrespective of their nationality or place of residence.

The completion of the Digital Single Market was identified as one of the ten political priorities of the EC [15]. By opening up digital opportunities for people and

[2] There has been some empirical studies conducted within the scope of the TOOP project however their focus were businesses rather than citizens.

business in the EU Member States, Europe's position as a world leader in the digital economy is aimed to be enhanced [15]. In order to streamline citizens' access to local authorities, the existing European portals need to be extended and linked to the Single Digital Gateway [15]. Currently, contact points between public authorities and citizens are fragmented and incomplete. Portals of the local authorities are mostly in the local language, which represent a substantial hurdle for cross-border activities.

By including this principle into the eGovernment Action Plan 2016–2020 [5], the EC aims to provide guidelines on supporting its implementation at regional and local levels. It is important to note that rather than developing a new portal, the Digital Single Gateway will be providing access to existing national portals from a single contact point. Overall, its development aims to support movement of individuals in cross-border settings by reducing the constraints imposed by existing borders.

2.2 The Once-Only Principle

OOP is one of the main cornerstones for enabling the efficiency of the Digital Single Market. The EC defines this principle as follows [5]:

> "Public administrations should ensure that citizens and businesses supply the same information only once to a public administration. Public administration offices take action if permitted to internally re-use this data, in due respect of data protection rules, so that no additional burden falls on citizens and businesses."

The particular benefits of the implementation includes reducing the administrative burden on citizens, achieving a more efficient government administration and increasing fraud prevention. According the SMART 2015/0062 report of the EC [16], the EU wide application of the OOP could result in annual net savings of as much as €5 billion per year. As a first step, the principle will be applied to exchange data between authorities within the same nation [6], as this is the prerequisite to cross-border data exchange. Implementation of it would likely require an update of the national infrastructures to ensure interoperability as well as a change in national legislations in particular data protection laws in some countries.

2.3 OOP Pilot Projects and Implementations

To explore and address the OOP related challenges in cross-border settings, the EC launched a call for proposals in 2016 [8]. After a careful analysis, two projects were selected for funding within the scope of the EU Horizon 2020 Research and Innovation Funding Programme: (1) *The Once-Only Principle Project* (TOOP) [17] and (2) *Stakeholder Community Once-Only Principle For Citizens* (SCOOP4C) [18].

The TOOP focuses on the application of the OOP for businesses, while the SCOOP4C has the focus of e-services for citizens. The TOOP has been subject to three pilot projects to explore the feasibility of the concept for businesses, while there has not been any pilot projects conducted within the scope of SCOOP4C. According to a recent position paper [14], the implementation of the OOP for businesses in the EU Member

States is still evolving[3], however not much is yet there at the cross-border e-Services level [16]. The EU strives to extend it across borders to further improve the efficiency of the Digital Single Market [15].

Another important issue is the clarification of which data regulation will be applied in case of cross-border e-services. Recognizing the importance of data protection in cross-border contexts, the EU has adopted a new data protection framework in 2016 to ensure data protection among the Member States. The GDPR came into force recently on May 25, 2018 [19]. Although the GDPR is aimed to supersede the EU Member State laws, changes in national legislations are necessary to ensure protection of citizens' privacy. Despite the required action of EU Member States, having a single regulation valid for all Member States is expected to be very beneficial considering the size of the EU in terms of eliminating conflicts such as which country's data protection regulation should be applied in cross-border settings [13].

Introduced as the most important change in data privacy regulation in 20 years [19], the data protection reform package underlines citizens' fundamental rights of data protection and foresees serious penalties up to 20 million Euro in case of breaches. Yet, in its current version, the GDPR is mainly focused on interactions between companies and individuals [19]. The implementation of the OOP would clearly benefit from a similar EU-wide data protection regulation defined for the G2C e-government context.

Estonia is one of the first countries that placed a special focus on the application of this concept both in national and in cross-border settings. The Nordic Institute for Interoperability Solutions Association was founded jointly by Finland and Estonia to develop online solutions to support cross-border operations as well as migration and commuting of citizens. In particular, data stored in numerous data repositories of the two countries are exchanged by utilizing the X-Road Technology. To support cross-border activities of citizens and businesses, its scope is planned to be extended to include the exchange of information between the tax authorities of the two countries [20].

Similar to its level of e-government take-off in the region [3, 4, 21, 22], Austria has taken the lead in implementation of the OOP concept in the DACH region. In Austria, financial aid is granted to families with newborn children automatically in most cases although up to six different government agencies are involved [18]. This is enabled through exchange of data in existing registries between public authorities. Due to positive feedback from the public, the Austrian government strives to fully automate the tax declaration and return service as well.

3 Data Analysis and Results

This section provides empirical research results to tackle the research questions posed in the introduction section. Four in-depth surveys were conducted online between 2014 and 2017 by using representative samples [3, 4, 21, 22]. The samples included Internet

[3] There are various implementations for businesses in EU Member States, which are beyond the scope of this paper and can be found in [14].

users in private households. The data is weighted to be representative of the online population by central features of gender, age and formal education. Three relevant questions of this research will be discussed in this paper. Surveys have included other questions such as the knowledge of and satisfaction with online public services, which will not be discussed here due to space limitations.

3.1 Adoption of G2C E-Government Services in the DACH Region

The analysis of G2C e-government adoption between 2014 and 2017 reveal that, e-Government adoption in the DACH Region has reached a stagnation point over the last few years (see Fig. 1). In all years of analysis, Austria had the highest take-off levels, followed by Switzerland. Germany, on the other hand, remains relatively behind, reaching to its lowest level in 2017. Overall, a significant rise in the e-government adoption rate could not be observed in any of the DACH region countries over the past few years, despite the advancements in IT technology and various national and EU wide initiatives.

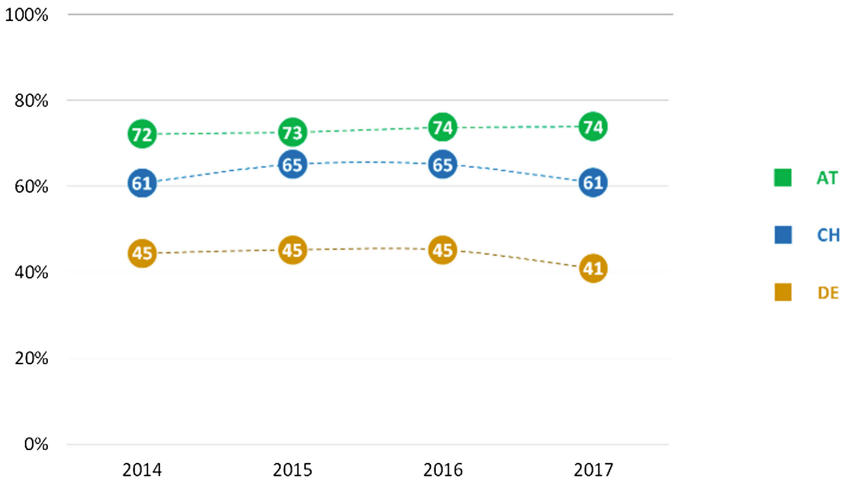

Fig. 1. Usage of G2C e-government services in the DACH region (in percentages, based on [3, 4, 21, 22])

These findings confirm the findings of other recent analyses on G2C e-government adoption. According to the Digital Economy and Society Index (DESI) of the EC [23], Austria was ranked 6th in 2016 and 5th in 2017 among 28 EU countries in terms of its Digital Public Services, being placed over the average rate of development within the EU. In the same analysis, Germany was ranked 20th in both years, resulting in placement below the average rate of development within the EU Nations. Switzerland was not subject to analysis in this research. Similarly, the Capgemini EU eGovernment Benchmark report [24] categorized only Austria from the DACH region as one of the best performing countries in terms of e-government.

3.2 Characteristics of a Modern Government Agency from the Perspective of Citizens in the DACH Region

Next, we have questioned the characteristics of a modern government agency from the perspective of citizens. This survey item has revealed insights on citizens' perceptions related to the national and cross-border usage of the OOP concept.

As Table 1 indicates, the top two characteristics are the same for all three countries. A modern public authority should provide quick response to requests (within 1–3 days). Allocation of online appointments with no waiting time at public authority is the second characteristic of modernity. Not surprisingly, these two characteristics are directly related to time savings, as traditional public authorities are widely accepted as being slow, inefficient and highly bureaucratic [25].

Table 1. Characteristics of a modern government agency from the perspective of citizens in the DACH region (based on [3])

	DE	AT	CH
	1,000	1,003	1,013
Quick response to requests (within 1–3 days)	67%	77%	74%
Online appointment allocation and no waiting time at public authority	63%	67%	62%
Availability of a central portal for citizens	47%	56%	47%
Availability of online information about the processing status of the application	46%	59%	50%
Continuous processing online	45%	58%	51%
Government request my personal data only once, which can be reused by other national authorities in compliance with data protection regulations	32%	46%	42%
Government requests my personal data only once, which can be reused by other European authorities in compliance with data protection regulations	14%	19%	14%

It is interesting to note that, the majority of the survey respondents did not consider the OOP as a characteristic of modern public authority. This perception was especially lower for German citizens compared to Austrian and Swiss ones. Exchange of personal data with other European authorities have encountered considerably much higher skepticism in all countries of analysis, although the compliance with data protection regulations was explicitly stated.

The discrepancy between perceptions toward personal data exchange between the national and EU-wide public authorities is striking. Although about one in every three citizens in Germany and one in every two citizens in Austria would consider exchange of personal data between national public authorities as a characteristic of modernity, more than four in every five individuals in all three countries have not seen any relation between modernity and exchange of their personal data at EU level.

The high sensibility of individuals regarding their personal data has also been subject to other studies in literature. Especially, the German nation is known to be

highly sensitive towards initiatives, which involve storing or transferring of personal data. Various initiatives in the past involving storage or transfer of sensitive personal data such as the Electronic Health Card Project have failed due to privacy concerns of citizens despite large amount of investments [26]. Such projects were heavily contested in the past by citizens, non-governmental organizations and political parties due to direct infringement to personal privacy [27]. This elevated sensitivity has a direct influence on adoption of such projects as well [28]. For instance, although one in every two German citizens (49%) has the new residency cards, less than one in every three of them (15%) decided to activate the eID function, which is the essential component for using the services online [3]. Therefore, even if some concepts such as the OOP are to be introduced at EU-level, their implementation strategy needs to be in each country separately because nations tend to have different levels of sensitivities. Such assessment should take the experience gained from similar projects in the past in consideration as well as a careful assessment of the sensitivities of the citizens which can be partially assessed by utilizing nationwide surveys and focus group analyses.

3.3 Opinions of Citizens Regarding Share of Their Personal Addresses in the DACH Region

The third question focused on understanding perceptions towards the implementation of the OOP concept in cross-border settings. In order to concretize the question, it was asked on the specific example of 'sharing their personal address with the public authorities of other EU Member States' (Fig. 2).

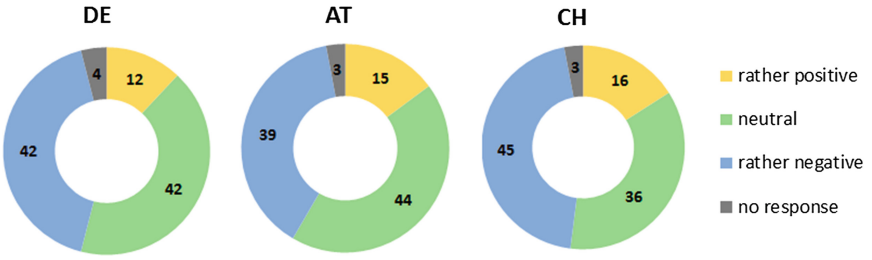

Fig. 2. Personal opinions of citizens in the DACH region towards sharing of their personal address with the public authorities of other EU Member States (in percentages, based on [3])

Similar to the previous question, this analysis has confirmed the reservation of respondents towards implementation of the OOP concept in cross-border settings. Only 12% of German respondents, 15% of Austrian respondents and 16% of Swiss respondents have perceived this approach as "rather positive". The percentages of "rather negative" responses were, in contrast, about two to three times much higher.

Overall, our research has confirmed the importance of data protection and privacy concerns within the G2C e-government context. Furthermore, we have clearly observed that the survey respondents were clearly much more reserved towards sharing data with

the governments of other EU Member States compared to data exchange between public authorities of their own country.

These results are in line with the findings of the recent TOOP survey, which revealed the perspective of businesses. Companies participating in the TOOP survey were significantly less willing to share data with authorities of other EU Member states compared to authorities in their own country [29]. Furthermore, data protection requirements were found as *the top challenge* for the pilot implementation of the TOOP project [29].

4 Discussion

Despite various national and EU-wide investments, G2C e-government adoption in the DACH region has reached a stagnation point. The OOP is one of the main concepts introduced by the European Commission to support modernization efforts within the EU. Citizens should deliver particular standard information to governments only-once and public authorities are allowed to exchange data among each other under consideration of the data protection regulations.

Yet, as being one of the initial studies reflecting the citizen perspective towards the implementation of the OOP, our research has shown that citizens of the DACH region are quite skeptical about it. Not even one every five respondents in the DACH region regarded exchange of their personal address with EU Member States as 'rather positive'. There is no doubt that this principle has a great potential of simplifying government processes, speeding up applications and related paperwork. This convenience comes, however, on the cost of data protection and privacy. Indeed, exchange of data between all EU Member States would increase the vulnerability of data enormously. Special caution is required as potential data breaches in the G2C e-government context is likely to have enormous consequences. For instance, citizens can update their credit card numbers in case of credit card fraud but sensitive personal data of individuals – such as the place of birth, data of birth or fingerprints – cannot be replaced [28].

Even though the data protection will probably to be ensured via EU-wide regulations, not every country has the most advanced security infrastructure in place. Even if they do, no country – including the ones enjoying highest e-government rankings – could be completely immune to data breaches and cyber-attacks. For instance, Estonia had to recently block all e-ID cards due to a massive security flaw, which could have irreversible consequences including identity theft [30].

Second, sharing data is not necessarily the interest of everyone. Not every citizen plans to study, work or move to another EU Member State in the foreseeable future. Thus, for such citizens, exchange of personal information with other EU Nations would only increase the vulnerability of their information and cause higher data protection and privacy concerns without any additional advantages. On the other hand, citizens go through various life events in their own country – such as getting married or applying for pension benefits – during which citizens have to interact with various national public agencies. This could explain the less skeptical perception of survey respondents towards exchange of information between national authorities as this could simplify the application as well as speed up the processing of it for citizens. The high number of

countries of the EU can be another reason of resistance towards EU wide data exchange.

Until now, there has not been much research in literature regarding the perceptions of citizens towards the EU wide implementation of the OOP concept. The findings of our study have confirmed that sharing personal data with public authorities is not necessarily seen as an aspect of modernity. In particular, citizens were relatively skeptical in case of cross-border data exchange. Furthermore, survey respondents would likely to have perceived the concept with even higher resistance, if the scope of exchange was not limited to their personal address. Yet, considering the amount of investments made by the EU, we assume that the scope of data exchange will go beyond the exchange of personal address to gain considerable benefits.

Despite importance of this issue, there has not yet been much research in literature on reflecting the citizen perspective and their perceptions on these issues. As the implementation has not started in most of the EU Member States, results of future empirical research can provide some valuable insights towards planning of implementations. As we have only considered the countries of the DACH region, results may not be generalizable to the remaining EU Member States. Therefore, we suggest future research to conduct further empirical studies to analyze the perceptions of citizens in other EU Member States.

To sum up, EC initiatives such as the OOP are quite promising; however, extreme caution is recommended regarding data protection. Every nation would likely have different levels of sensitivity, which needs to be taken into consideration. In particular, benefits added by increased convenience provided by data sharing should be carefully analyzed against the increased risks of data protection and its potential implications. The rising cyber criminality, massive data protection flaws and the overall vulnerabilities in online systems are frequently discussed in various media. E-government is unfortunately no exception. The European Commission and the EU Member States should explicitly state which personal data would be exchanged, under which circumstances as well as the rights of citizens such as the necessity of consent and its withdrawal at any time. After all, reaching the annual net savings of as much as 5 billion per year would only be possible, if its implementation is planned and conducted with the utmost care.

References

1. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: an integrated model. MIS Q. **27**(1), 51–90 (2003)
2. Beldad, A., De Jong, M., Steehouder, M.: How shall i trust the faceless and the intangible? A literature review on the antecedents of online trust. Comput. Hum. Behav. **26**(5), 857–869 (2010)
3. Krcmar, H., Akkaya, C., Müller, L.-S., Dietrich, S., Boberach, M., Exel, S.: eGovernment MONITOR 2017: Nutzung und Akzeptanz digitaler Verwaltungsangebote - Deutschland, Österreich und Schweiz im Vergleich. Initiative D21, fortiss GmbH (2017). http://www.egovernment-monitor.de/fileadmin/uploads/Studien/eGovMon2017_RZ_FINAL_WEB_NEW.pdf. Accessed 20 June 2018

4. Krcmar, H., et al.: eGovernment MONITOR 2016 Nutzung und Akzeptanz digitaler Verwaltungsangebote - Deutschland, Österreich und Schweiz im Vergleich. Initiative D21 e. V., Institute for Public Information Management (ipima) (2016). http://www.egovernment-monitor.de/fileadmin/uploads/Studien/2016/160915_eGovMon2016_WEB.pdf. Accessed 20 June 2018
5. European Commission: EU eGovernment Action Plan 2016–2020 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0179&from=EN. Accessed 20 June 2018
6. European Commission: Tallinn Declaration on eGovernment (2017). http://ec.europa.eu/newsroom/document.cfm?doc_id=47559. Accessed 18 June 2018
7. Deutscher Landkreistag: The New German Coalition Agreement (2018)
8. Wimmer, M.A., Tambouris, E., Krimmer, R., Gil-Garcia, J.R., Chatfield, A.T.: Once only principle: benefits, barriers and next steps. In: Proceedings of the 18th Annual International Conference on Digital Government Research Staten Island. ACM, New York (2017)
9. Krimmer, R., Kalvet, T., Toots, M., Cepilovs, A., Tambouris, E.: Exploring and demonstrating the once-only principle: a European perspective. ACM (2017)
10. Buyle, R., et al.: OSLO: open standards for linked organizations. ACM (2016)
11. Veiga, L., Janowski, T., Barbosa, L.S.: Digital government and administrative burden reduction. ACM (2016)
12. Kalvet, T., Toots, M., Krimmer, R.: Contributing to a digital single market for Europe: barriers and drivers of an EU-wide once-only principle. In: Proceedings of the 19th Annual International Conference on Digital Government Research Delft, Netherlands. ACM (2018)
13. Tikhomirova, A.: Reinforcing trust and security in digital services and in the handling of personal data. InterEULawEast: J. Int. Eur. Law Econ. Market Integr. **3**(1), 145–153 (2016)
14. Krimmer, R., Kalvet, T., Toots, M., Cepilovs, A.: The once-only principle project: position paper on definition of OOP and situation in Europe (updated version) (2017)
15. European Commission: A Digital Single Market Strategy for Europe (2015). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0192&from=EN. Accessed 20 June 2018
16. Cave, J., Botterman, M., Cavallini, S., Volpe, M.: SMART 2015/0062: EU-wide digital Once-Only Principle for citizens and businesses (2017). https://ec.europa.eu/esf/transnationality/filedepot_download/1671/1692. Accessed 15 June 2018
17. Krimmer, R.: TTOP Webpage (2017). http://www.toop.eu/. Accessed 01 Feb 2018
18. SCOOP4C Webpage: Austrian federal government commits to implementing once-only principle in new e-government plan. https://www.scoop4c.eu/news/austrian-federal-government-commits-implementing-once-only-principle-new-e-government-plan. Accessed 20 Feb 2018
19. European Commission: General Data Protection Regulation (GDPR) Portal (2017). https://www.eugdpr.org/. Accessed 15 June 2018
20. Nordic Institute for Interoperability Solutions (NIIS): Digital society solutions and cross-border cooperation. https://www.niis.org. Accessed 15 May 2018
21. Krcmar, H., Wolf, P., Rau, L., Till-Stavrakakis, V.: eGovernment MONITOR 2014 Nutzung und Akzeptanz von elektronischen Bürgerdiensten im internationalen Vergleich. Initiative D21 e.V., Institute for Public Information Management (ipima) (2014). http://www.egovernment-monitor.de/fileadmin/uploads/Studien/2014/eGovMon2014_web.pdf. Accessed 15 June 2018

22. Krcmar, H., Wolf, P., Zepic, R., Müller, L.-S., Till-Stavrakakis, V., Boberach, M.: eGovernment MONITOR 2015 Nutzung und Akzeptanz von elektronischen Bürgerdiensten im internationalen Vergleich. Initiative D21 e.V., Institute for Public Information Management (ipima) (2015). http://www.egovernment-monitor.de/fileadmin/uploads/Studien/2015/150715_eGovMon2015_FREIGABE_Druckversion1.pdf. Accessed 15 June 2018
23. European Commission: The Digital Economy and Society Index (DESI) https://ec.europa.eu/digital-single-market/en/desi. Accessed 10 June 2018
24. Capgemini: eGovernment Benchmark (2017). https://www.capgemini.com/wp-content/uploads/2017/11/2017-egovernment-benchmark_background_v7.pdf. Accessed 10 June 2018
25. Lotze, B.: KVR: Lange Schlangen und kein Ende, Süddeutsche Zeitung (2018). http://www.sueddeutsche.de/muenchen/behoerden-kvr-lange-warteschlangen-und-kein-ende-1.3993214. Accessed 16 June 2018
26. Elektronische Gesunheitskarte offenbar vor dem Aus (2017). http://www.sueddeutsche.de/wirtschaft/e-card-elektronische-gesundheitskarte-offenbar-vor-dem-aus-1.3617842. Accessed 10 June 2018
27. Akkaya, C., Wolf, P., Krcmar, H.: Factors influencing citizen adoption of e-Government services: a cross-cultural comparison (Research in progress). In: Proceedings of the 45th Hawaii International Conference on System Science Hawaii, USA. IEEE (2012)
28. Akkaya, C.: A comprehensive analysis on citizen adoption of e-government services: a cross-cultural analysis. Technical University of Munich (2016)
29. Kalvet, T., Toots, M., Krimmer, R.: The once-only principle project: drivers and barriers for OOP (1st version) (2017). http://www.toop.eu/sites/default/files/D27_Drivers_and_Barriers.pdf. Accessed 15 June 2018
30. Moon, M.: Estonia freezes residen ID cards due to security flaw (2017). https://www.engadget.com/2017/11/04/estonia-freezes-resident-id-cards-security-flaw/. Accessed 01 June 2018

Open Data, Linked Data, and Semantic Web

Investigating Open Government Data Barriers

A Literature Review and Conceptualization

Jonathan Crusoe(✉) and Ulf Melin

Division of Information Systems, Linköping University,
581 83 Linköping, Sweden
{Jonathan.crusoe,Ulf.melin}@liu.se

Abstract. When focusing on open government data (OGD) publishing and related barriers, there are several complexities present. Largely, current research is focused on publishing and usage of OGD; and we argue that there are a need to investigate and to systematise OGD barrier research in order to understand and outline an expanded scope of the phenomenon. We expand by clarifying barriers linked to the release decision and the data's organisational context. To investigate the OGD barriers, we conduct a systematic literature review, identifying 34 articles as a point of departure for our analysis. From these articles we create, present and discuss illustrations on historical development, barrier types, and different research focuses on OGD. When analysing the articles, we identify a focus on technical, organisational, and legal barrier types, while studies on open data usage and systems are less frequent. Our analysis also identifies some possible open data research barriers. In the article we also relate barriers to an expanded OGD process (Suitability, Release, Publish, Use, and Evaluation), identifying 46 barriers with possible linkages. The results is an expanded scope and a conceptual illustration of OGD barriers.

Keywords: Open data · Open government data · OGD · Barriers
Risks · Challenges · Impediments · Myths · Process · Literature review

1 Introduction

Open government data (OGD) refer to data sets that government agencies make available for third-party usage. Putting OGD in contrast to history, where public organisations produce data enclosed in silos where its value is limited. One purpose is to make data accessible to a broader audience by stimulating the third-party development of new e-services (e.g. apps for weather forecasts and smarter travelling routes). By opening data for others to reuse, we gain OGD [17]. OGD is surrounded by many expectations of service innovation, increased transparency and inclusive government agencies [3].

As a concept and phenomenon OGD is promising, but at the same time challenging. OGD research and practice has shown several complexities when publishing data [e.g. 3, 9, 12, 17, 39, 42]. Challenges, or barriers as we conceptualise them in this

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 169–183, 2018.
https://doi.org/10.1007/978-3-319-98690-6_15

paper, are described under different conceptual labels in previous research, such as; barriers [e.g. 4], impediments [e.g. 42], myths [e.g. 13, 20], and challenges [e.g. 28]. Multiple analytical perspectives are also present in previous research; focusing on barriers linked to release [e.g. 4], adoption [e.g. 20], cognitive [e.g. 37], agenda [e.g. 26], process [e.g. 9], and use [e.g. 42], use and publish [e.g. 5]. OGD is also placed in different larger context, by using, e.g. an ecosystem perspective [12]. In present OGD research, there is a tendency to study specific parts or specific phases of a larger system or process.

Another complexity is the nature of the OGD barriers and the OGD process. The barriers can e.g. be linked to different process phases, and there are several ways to divide the OGD process into phases [e.g. 3, 7, 12, 42]. Which is also further complicated when different actors can participate in one or several phases [42], and activities in one phase can create barriers in a later for someone else [39]. The complexity can be visualised as a barrier network [18], which include a reversal of the cause-and-effect; a barrier can come from a later solution. We, therefore, see earlier OGD research giving room for specialised solutions for unique and challenging problems; where solutions can cause barriers in both directions of the OGD process. There is a need to understand and systematise existing OGD barrier research with an expanded scope.

We expand the scope by taking a step back and view other actors and activities surrounding the publisher and open data publication, which includes more OGD barriers. We investigate the OGD barriers in earlier research with a systematic literature review and expand on earlier works, such as Zuiderwijk and Janssen's OGD process [38], through a systematisation. An expanded scope can serve as inspiration for a future research agenda and practice in the field. The result illustrates insights into the complexities and challenges encountered when publishing. We also outline some future OGD barrier research avenues. The following research questions are focused in this paper:

- What OGD barriers have been identified in previous research?
- Where are the OGD barriers encountered in the OGD process and how are they connected?

In the following sections, we will start by describing our research approach. Then we will discuss the identified barriers from the literature and chart illustrations, and in the next section presents the OGD barrier systematisation. We end the paper with a conclusion containing limitations, future research, and implication.

2 Research Approach

We conducted a systematic literature review [3, 10, 17] where we analysed the findings through coding [32] to make illustrations and a systematisation. The literature review [3, 10, 17] started by defining keywords, and then selected a database that was searched, results were filtered and summarised, and lastly, we analysed the findings. We created simple keywords by combining "open data" with "barrier", "risk", "challenge", or "impediment". For database, we selected Google Scholar as it has good coverage,

with recall and precision that is above average for simple keyword searches [36]. For each keyword search, we looked at the first 50 results. We first searched all years, but also wanted to be sure that we caught the latest and relevant articles, we therefore explicitly searched the years 2016 and 2017. This gave us a total of 600 articles; we discarded duplicates. We identified relevant articles by studying the title and Google Scholar summary for concepts relating to our keywords. Relevant articles were summarised with article id, authors, title, year, publish origin, topic, and important conclusions, if needed, purpose and method were included, such as the case of the systematic literature reviews. Later when analysing the articles, we discovered interesting references and snowballed. We summarised their content in the same way as above. This resulted in a final set of 34 articles. We analysed the articles to present them as data (see Sect. 3) and to systematise OGD barriers (see Sect. 4). For the data presentation, we decided to analyse development, types, and focuses, making a distinct coding effort for each into a representative illustration. When coding for development, we sorted the articles based on publication year. We then used them to form a graph and coded based on what and possible why. For analyzing the types, we divided articles into two groups: specific articles (containing a thematic division or typology of barriers into types), while the general articles contain barriers (but not structured by type, e.g. [5] structured by role, or [4] structured by type). We coded the first group by following their thematic division or typologies. After that, as part of categorization, we integrated the second group into the first through coding. When analysing focus we assumed that the abstracts provided an overview of the work with information about the background, objective, methods, results, and conclusions [43, 44]. We therefore initially coded [32] the abstract for research focus. If the abstract lacked the necessary information, the whole paper were explored. After codes were categorised into groups and we merged and split groups to achieve a pedagogic presentation. The product was a sketch that we as researchers discussed and analysed in a generative dialogue, and some final changes were made. The results were the three illustrations in Sect. 3. For the systematisation, we first identified an OGD process, with different parts and activities. We then categorise coded [32] the 34 articles into it. Giving the barriers a chance to expand the OGD process by e.g. including concepts. After, we identified possible linkages between barriers following the network idea of Huang, et al. [18]. Linkages were identified when barriers shared common core phenomena or phenomena were dependent on each other. As a final step, we refined the categories by reflecting on their meaning.

3 The Identified OGD Barrier Literature

Below, we present the articles as data through illustrations following our literature analysis, in the following order: historical development of the OGD barrier discussion, barrier types, and research focuses. We note that some research only acted as *enablers* [17, 21, 35] (used to snowball) and was, therefore, not used in the later systematisation.

3.1 Historical Development

In this section, we present the historical development of the OGD barrier discussion. In Fig. 1, we have divided the number of articles published 2011–2017. In 2014, we identified an increase of published articles. We will use this peak as an important point and discuss the time-period before and after.

Before 2014, we have a total of 4 articles researching OGD barriers. In 2012, we identified what seems to be the start of a crystallisation that had its full expression in 2014. E.g. the literature review [42] focused on the user OGD barriers. The authors pull together research from different research domains. Suggesting that OGD challenges encountered covers a variety of domains and around 2012 these started to crystallise into concepts such as open data barriers, challenges, myths, risks, and impediments giving way for a collective discussion.

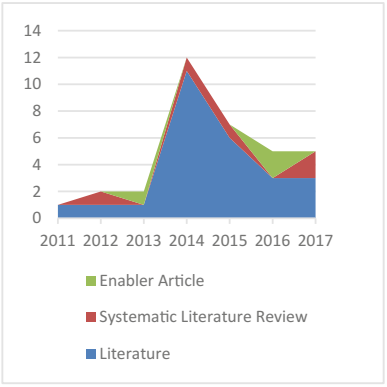

Fig. 1. Historical development

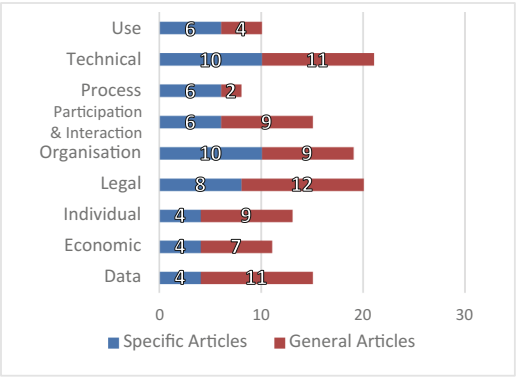

Fig. 2. Barrier types found in the literature

After 2014, we identified 5 articles researching OGD barriers per year, a decrease compared to the 12 articles in 2014, conflicting with what is known in 2017: open data barriers are evident [e.g. 5, 18], something that is not reflected in the research attention. When analysing the articles, we identified a pattern. Myths, for example, [13, 20, 28] are not reused or connected to other OGD barrier research. Research has not revisited the focus on implementation and use from a multi-level perspective [26]. A process perspective [e.g. 3, 39] seems to be useful and often comes with a focus on the publisher, and therefore, lacks the barriers existing on the "outside" between publishing and feedback. Such as barriers for hackathon apps to reach the market [14]. We see a focus to understand barriers and their linked parts, leading to rather limited attempts to solve complex issues from multiple directions; from the micro to the macro perspective.

When looking at probable causes for the decline above, it can be that research is in an incubation period or have expanded. If an incubation period is present, we identify two OGD research barriers; (1) we are still searching for the abstract approach to handle barriers, e.g. like outlined in [4] and [18]. (2) We have not crystallised our

knowledge about barriers, meaning that we are still working towards comprehension; one attempt has been made by [5]. Both fit with the pattern above. On the other hand, if research has expanded; it can have expanded into new domains. E.g. from government into innovation [e.g. 14]. With this, the label used to address the OGD barriers might have changed to fit the new domain. Fitting with what we identified in 2012. There is also a possibility that the research has moved over to solutions. For both, we see two accompanying OGD research barriers; (3) keeping the knowledge body together between domains and (4) researching solutions that will not lead to barriers in other parts of the OGD process. To conclude, we identify a complexity in research and practice covering the micro technical and social (formats, metadata, skills, knowledge, etc.) and the macro technical and social (internet, ecosystem, society, etc.) making open data barriers a complex and challenging research task, as it covers multiple domains where barriers exist outside a particular actor.

3.2 Barrier Types

In Fig. 2, we present barrier types identified in the 34 articles. We have coloured the two groups mentioned in Sect. 2. When studying the different barrier types, it was common to find technical, organisational, and legal barriers (identified in 21, 19, and respectively 20 articles), something that was not surprising as open data originates in the publisher organisation's social and technical systems [3, 9, 39]. On the other hand, data barriers appeared in 15 articles, but were sometimes fused with technical barriers such as format and quality; they tended to be linked. Making data barriers implicate technical barriers. For open data, the value is generated from reuse [17] and we, therefore, expected use to be equally studied as technical, organisational, and legal barriers, it was identified in 10 articles. However, use was sometimes fused with technical barriers through access and findability. Use was also often fused with participation and seemed to be assumed to go together, as they are often integrated into the OGD processes [e.g. 3]. Making use through implicitly equally studied to technical, organisational, and legal barriers. Equally important, but less studied was the OGD process (found in 8 articles), where there were a gap between accessing data and feedback, where usage or application barriers were absent. We see that there is a lack of a wider scope when studying OGD barriers. In the OGD process for the publish-and-use relationship, we found a tendency for barriers to affect both providers and users while not stated as such. Use, participation, and process barriers need study beyond the direct interaction between providers and users. Largely, it remains unknown how organisations [e.g. 9] connect with publisher-user relation [e.g. 40] and affects the user [e.g. 6] and how this, in turn, affect participation and the organisation as a publisher. Turning our attention towards the more individual-oriented barriers, such as skills, perceived risks, knowledge, and awareness, they have been studied in 13 articles, e.g. in the role of the OGD user. These barriers differ from the above as they focus on how a person needs to adapt to open data, rather than changes in techniques and organisations. As individuals populate organisations, it can be challenging to separate the two, but we think that, based on the patterns above, it is relevant to conclude that we should aim to focus more on individual-oriented barriers in organisational settings. The final barrier type is economic, which were identified in 11 articles; we found discussions about

costs, but no clear evidence that benefits would outweigh them. There was also a lack of connection between costs, organisations, and solutions (both implementation and maintenance) to barriers. As pointed out by [17] data generation and maintenance cost money. Articles tend to focus less on the cost associated with solving barriers and challenges. However, it is not the case that costs are ignored, e.g. [5] brings up cost issues. Two major issues seem to be calculating implementation costs for overcoming barriers and maintenance costs versus benefits, e.g. how do we calculate the value of increased transparency, democratic accountability, and stimulation of innovation [20]. We see a need for more research about economic barriers in comparison to other barriers.

3.3 Research Focuses

In Fig. 3, we present the identified research focuses sorted into categories. Each study is a box with citation number to the left and to the right; we have at the top focus and publishing year below. Underlined numbers refer to systematic literature reviews. We categorised the studies, which resulted in the following categorical descriptions: Assumption studies explore barriers steaming from assumptions. Decision studies approach barriers about considerations; such as the decision to publish or not. Interaction studies study barriers in the relation between actors. Legal and privacy studies how law and norms interfere with publishing. Publishing studies focus on barriers related to the publishing process. Review studies are exploring the literature. Society studies investigate open data in society. A system perspective on barriers is also present. Usage research study barriers that impede open data user reuse.

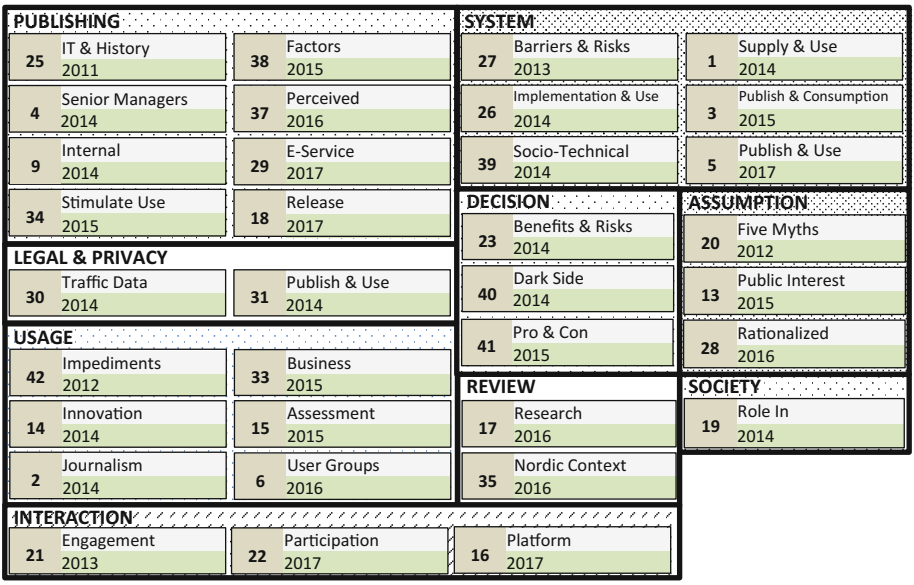

Fig. 3. Research focuses found in the literature (Based on abstracts).

In Fig. 3, we see an ongoing focus on publishing, usage, and systems over the years. The strongest focus was on publishing, while both usage and a system approach came second. Together with frequency, technical, organisational, and legal barriers are mentioned in favour overuse and participation (see Fig. 2), we have identified that challenges seem to be inside public organisations (or we need more research on use and usage). Looking at the figure, we see that interaction was studied under the years 2013 and 2017, with focuses on engagement, participation, and platform. In the earlier section, use barriers were often fused with participation and interaction barriers, where use seems to link to publishing through technical barriers. We, therefore, expected to see a balance amongst the barriers between users and publishers, with some focus on equalisation of expectations and division of labour [28], but instead found a focus on how publishers "imped" interaction where users were less problematized. For studies focusing on usage, they focused either on certain user groups and their special usage [e.g. 2, 16] or specifically on the use of provided data from a user perspective [e.g. 15, 42]. Here we found a gap in how different use barriers connect to different usages and users (see Davies [11] for examples of usages). The relative low societal focus can reflect the general maturity of the OGD field, being rather early in its lifecycle. We have, in research and practice, not yet solved issues surrounding publishing, using, and interaction. In the section above, legal barriers were mentioned second to most, but here we see that it has been studied the least. Legal and privacy barriers seem to accompany other focuses. It can be that they are solved by practical work, rather than theoretical. There is a lack of research on assumptions and publish decision barriers, unpacking this could lead to further insights about the publisher.

4 Systematization of Open Government Data Barriers

In the following section, we expand the scope surrounding the OGD process and systematise the OGD barriers. The OGD process consists of phases that involve data generation, data publication, data discovery, data usage, and user feedback [3, 39], which is the final process after data publishing. Too, understand the barriers behind publication we take a step back. Open data originates from an organisation where it starts as closed, which means that the data has an organisational context [9], before and after publication. In between data being close and open, there needs to be a decision for data release [23, 40, 41]. Meaning that we have five distinct processes containing barriers; when (1) identifying data's suitability for publishing, (2) deciding to release, (3) publishing the data, (4) someone uses the data, and (5) evaluating the impact and collect feedback. The processes are further complicated when it is known that activities in earlier phases can cause barrier further down [39] and connect with each other [18], which we argue also applies to our five distinct processes.

The relationship between the distinct processes is complex. Placing the processes beside the ecosystem [12], we find that they are connecting through ecosystem elements. We view elements as a complex network of participants with their challenges (some not related to OGD). The identification of suitable data starts when there is interest in some data. The suitable data is input into the other processes. Inside a decision network, we find decisions to release processes. Participants are the

organisation's internal management and enthusiasts, but also outsiders, such as activists, researchers, and politicians [e.g. 28]. If the decision is release, the publishing process triggers inside an organisation. The publisher (people behind data publishing) will then work to publish the data into the data market (people behind data use and products [e.g. 24]). Where users can transform the data through use into societal impact (people benefiting, affected by, or using open data outside or at the edge of the data market) [24]. Impacts could be of political and social, economic, and operational and technical nature [20]. The decision network then evaluates the impacts, which can concern the publisher, users, and citizens. Evaluation can happen through feedback, discussion, [39] and some evaluation is internal or external to the OGD process [12]. The processes form a cycle starting and ending with the decision network. For an example of the cycle; low-quality data negatively affects usage that results in an impact on the perceived usefulness of open data [17], resulting in less reason to publish. Low usefulness may lead to a revaluation of the open data's suitability. In the following subsections, we will present barriers identified for each of the five distinct processes. Bold words are barrier **categories**, while cursive are *coded*; behind both, there are citations to show quantity and origin.

4.1 Identifying Data Suitability

In the process of identifying data's suitability as open data, barriers will be encountered in the relationship between the data and organisation. Here the first thing to study is the relationship towards the *Core Task* [5, 9, 19, 25, 39, 42]. We need to explore what role open data has in the organisation and under what time-period. As been noted by Jaakkola, Mäkinen and Eteläaho; *"Some general open data related problems and challenges are…[a] guarantee of maintenance and updatedness of data is missing"* [19, p. 33]. One question is if the organisation have an interest to continually and actively collect the data. *Data* [3, 15, 20, 42] is produced by the organisation and will, therefore, have a meaning dependent on that context, with accompanying assumptions. We will have to study the data's properties (e.g. static or dynamic), context (e.g. language), and quality (e.g. completeness, fragmentation, accuracy). The production of the data happens in the *Collection Process* [3, 9, 20, 27, 42] of the organisation, which can be integrated into other activities. In it, data is collected, created, or generated. Tools, filters, data processing and methods are used, which can create barriers for later publishing and reuse. E.g. filtering before storage can result in *"No access to the original data (only processed data)"* [20, p. 262]. The *Storage System* [3, 5, 9, 18–20, 22, 25, 26, 38, 39, 42] stores data for the long term. The form and location of the storage can vary; with everything from paper-based to digital and in-house to the cloud. At the same time, the owner of the data and the storage can differ. Barriers can be an inability to convert data, improve storage system, or the share amount. Something that brings us over to *Path Dependency* [20, 25, 28, 42] as a source of barriers. E.g. municipalities seldom considered document formats when buying software, instead focusing more on specific office applications [25]. Meaning that they might have bought a system based on closed formats that are hard to open. In relation to collection, storage, and core task we have the *Internal Usage* [9, 19] that will form the data with or for a purpose. A forming that might make the data unsuitable for external reuse. The

data's **Suitability** [9, 26, 38, 39] as open data, comes from itself and relationship with its organisation. We can, therefore, talk about suitability barriers that complicate publication and reuse, where solutions can problematize the organisation's day-to-day activities.

4.2 Decisions to Release

Once suitable data have been identified, a release decision can be made. The organisation might have several datasets and no idea which to publish, which relate to **Market** barriers. Here *Supply and Demand* [26, 27, 40] can identify sought-after data; if demand is neither expressed or heard then it forms a barrier as an organisation cannot prioritise publishing. Another side is *Competition* [3, 31]. Outside the organisation there might be actors competing with or around the data. Releasing the data can dislodge their business models. On the other hand, if we turn our attention towards the publisher we have **Organizational** barriers. If publishing the data do not align with the organisation's *Objectives* [20, 34, 41], this may form a barrier as other tasks are prioritised. Prioritization can also be affected by the availability of resources. Which makes *Cost and Income* [3–5, 9, 16, 20, 26–29] interesting, as it is known that open data can lower costs, but also income [4, 5]. Meaning that open data has a risk to threaten resources used in core tasks. This also moves us to the *Division of Labour and Alternatives* [28, 29]. For Division of Labour it is a question about expectations on publisher and user; e.g., if there is a need, should publisher or user implement visualisation tools. Placing this alongside objectives and resources, it also becomes a question about alternatives. An alternative to open data within a government organisation, can, e.g. be to develop and launch e-services [e.g. 29] that are prepacked and controlled in another way that is publishing OGD for any user and any purpose. Another potential for declining to publish are **Consequence** barriers. Here *Regulation and Legislation* [3–5, 19, 20, 22, 23, 27–29, 31, 37, 40, 42] can block or problematize open data publishing. As an example, [4] identified *"Irish law can be a considerable barrier to making data available."* (p. 145). In relation to the laws we also have *External Safety* [4, 23, 29], where [23] tells us *"Detailed data about infrastructure (power plants, dams, transmitters etc.) might be misused to cause damage to the infrastructure."* (p. 35). If something does go wrong (abuse, misinformation, fraud, accidents, etc.) there will be the question of *Liability* [3, 4, 20, 23, 29, 31, 39, 40]; if this is unclear, both publisher and users might be at risk of unexpected and unwanted responsibilities. Laws and norms also make *Privacy* [1, 3, 5, 16, 18–20, 23, 26–31, 39–42] intrusions something that must be avoided (a barrier), where the solution might be to clean the data, which can also make it lose value. Another solution is to use *Embargo Periods* [39–41]. The final source of barriers is a lack of ***Skills for Release*** [26]; the organisation might lack the skills, competencies, and tools to support release decisions. If pushed, they might publish data that is a threat to external safety, useless, or reveals private information about citizens.

4.3 Publishing the Data

Publishing data is no easy task as there will be internal barriers, but also a need to consider usage. One barrier that can cause future barriers is a lack of ***Skills for Publishing*** [1, 3–5, 15, 18, 20, 22, 27, 34, 38, 39, 42], as it can e.g. result in inadequate infrastructure performance, lacking metadata descriptions, or inaccessible data. Another hurdle is **Service** barriers. *Ownership* [5, 9, 23, 27, 31, 34, 39–41], as Susha et al. [34] tells us *"For Groningen municipality, the biggest problem was to convince the data owners to publish data which they considered were not of sufficient quality."* (p. 193); can be a barrier when the organisation are unsure of service quality. Unclear data ownership can also make it hard to know what can be published. Sufficient *Quality* [1, 5, 15, 18–20, 22, 23, 26, 39–42] must also be sought and guaranteed; else it can lead to future use barriers. The data might be the wrong format, unreliable, inaccurate, and incomplete. Quality connects technical infrastructure, data's suitability, and the users' needs. For the technical infrastructure, we are interested in external and internal barrier sources. For external barriers, we must consider **Opening** barriers. Here *Access* [3, 5, 15, 20, 27, 34, 39, 41, 42], *Metadata* [3, 5, 18, 20, 26, 27, 34, 38, 39, 41, 42], and *Format* [1, 3–5, 15, 18, 20, 22, 27, 34, 38, 39, 42] can cause problems for prospecting users. Paywalls, registrations, and API only downloadable data can form access barriers. While metadata barriers can be none explained context, language, or purpose for the data, which can make it hard to identify opportunities and usages. On the other hand, formats that are badly documented or proprietary can hinder data integration and long-term sustainable usage. For the publisher selecting a format when there is no standard can be a barrier. The original storage system format can be a problem too, as exampled by Kassen [22] *"…some data are really challenging to publish in a computer-readable format due to the fact that they only exists as a paper document…"* (p. 317). In addition, converting the storage systems format to one that is open, machine-readable format can also be a barrier. For the internal technical infrastructure, we have **Administration** barriers, which likely will be experienced once data is published. The first is *Performance* [3, 4, 16, 23, 28, 29] where infrastructure might not be able to handle the external requests, which can affect the organisations day-to-day activities, but also impact users' perception of data quality (e.g. overloads that leads to system crashes). The second is *Maintenance* [5, 16, 18–20, 27, 29, 39, 42], where a possible increase in requests might require upgrading the technical infrastructure but can also lead to increased "wear and tear" (an increase in hardware replacement). As publishing is about opening closed data to third-party reuse the organisation is also opening for the public, therefore, the final is *Security* [5, 20, 37]. Where lacking security can lead to a threat against the core task, but also the external safety, as devious users gain access to private information or have the ability to damage the organisation's systems.

4.4 Using the Data

While there are user barriers outside the control of the publisher, some can be mitigated. First users must be able to find the data. **Findability** [3, 5, 16, 18–20, 23, 39, 41, 42] refers to the ability of the user to discover and identify the data, which to a degree can be solved with an open data portal. Findability related barriers could come from

different sources, such as a market flooded with similar datasets or diffuse names. Zuiderwijk and Janssen [39] tells us *"Making open data findable poses the challenge for different government organizations to collaborate."* (p. 122). Meaning that findability barriers is in the relation between publisher and user, between publishers, and both actors relation to the open data portal. Once users have found the data, they want access and read the *License* [3, 5, 14, 15, 19, 27, 31, 39, 41]. The license can cause barriers when not explained, incomprehensible, or complicated. Another part is if the license restricts usage to much or is incompatible with other licenses. There is also a question about user *Responsibility* [5, 14, 31, 40, 42] that connects back to liability. An example is *"…the UK Open Government Licence for instance requires that the re-user does not mislead others or misrepresent the Information or its source [45]…"* [31, p. 7]. The authors argue that such clauses should not be needed, but also admit that it seemed to make public agencies more comfortable with publishing. We see that an unclear license can be a barrier for users, while unclear user responsibilities can be a barrier for the publishers. At this stage, the user might want to read *Documentation* [5, 15, 39] that explains how the data can be used, accessed through an API, or see processing examples. Documentation connects back to metadata, but rather than explaining the what, it supports the user in the endeavor to use. In common for all above we have *Language* [5, 27, 39, 42] as a barrier. If the users cannot read or understand websites, licenses, documentation, or the data it will likely complicate usage. We also have the *Paywall* [5, 15, 20, 39, 42]; the user must pay for data. A paywall can stop the users, especially if there is no way to express feedback of or experiment with the data. This means that a paywall is a barrier that can worsen if there is also bad metadata or documentation. Users will also experience barriers if either license, format, IT-systems, or data content lacks *Interoperability* [19, 23, 26, 34, 42] with other data. The ability to combine data is important for open data to reach value. Interoperability can also come with collaboration barriers for publishers as standards are implemented. Another barrier is *Timely* [39–41]; the data needs to reach the user before it spoils. The barrier here is long embargo period or slow access. While mentioned barriers focuses on the experiences of the users, we can also look at user **Diversity** barriers. *Different Users* [2, 6, 14, 18, 33, 39, 42] comes with diverse needs and goals. E.g. visualise data or integrate data into an app. If not meet, users might be unable to use the data. On the other hand, publisher can find it hard to identify and meet users' needs. *Tools/Software* [3, 5, 15, 20, 34, 39, 42] can help users, but then we stand in for the question of who should implement them and how far. We also have the other side where data might need special tools to use, which can be a barrier if they are unusable or cost. Users also have their own barriers, for example *"… current challenge facing data journalism in Sweden is the lack of time to collect, analyze and present data. This may be a result of streamlined organizations as well as not having enough competence and skills to efficiently work with data."* [2, p. 404]. If data require special *Skill for Usage* [5, 6, 13, 14, 16, 18, 20, 27, 34, 39, 42], users without them are at a disadvantage and how to supply these will be a barrier for both publishers and users. This leads us to **Usability** [3, 5, 16, 18–20, 29, 34, 39] as a barrier. Data loses usability when users face some combination of the barriers above or cannot use the data, in addition high amounts of unusable data in the data market can lower the findability of high quality data. This means that a combination of usability, awareness, and findability can result in *Low Usage* [14, 15, 34, 39].

4.5 Evaluation

Once data is published and, hopefully, in use, it is relevant to evaluate the impact and collect feedback to improve. Here we can study **Community** barriers, such as lacking *Channels* [5, 20, 22, 34, 42] used to collect and communicate and low open data *Awareness* [3, 18, 22, 31, 34]. If end-users are not aware of their open data usage, they cannot tell how it impacts. In the inverse, if users have feedback, but no channel, the process will be hard to improve (and we lose impact insight too). Users that are not aware of OGD and its possibilities will not show *Interest and Need* [1, 22, 26, 27]. Even if they have awareness and channels, they can express interest, to express need they require access to the organisation's metadata inventory. Here we have connected interest and needed to supply and demand, as without expressed interest and need it will be hard to evaluate demand for some closed data. As we argued above, different users have diverse needs, and it, therefore, becomes important to think about *Participation* [3, 13, 16, 34] in evaluation and feedback. Not involving a diverse group of users can result in a specialised solution that creates complications for other users. There is also another barrier related to participation, as Hellberg and Hedström [13] say *"We believe that not everyone is interested in using public data, even if they have the necessary resources and competences."* (p. 47). We see a possibility that low awareness and uniform participation can lead to specialised solutions, where not all with capability are interested in participation, which lowers usability by exclusion. Let us say we have participants that want to give us feedback; then we need to think about **Coordination** barriers. Here *Quality in Evaluation* [15, 20, 22] can be a challenge, as feedback can be in multitude and varied or users are without supporting documentation and tools; the user might also not know what to look for. Once we have participants with feedback, it becomes a question of *Feedback and Error Handling* [34], which connects back to administration. If errors and feedback are not handled problems will continue constraining the open data impact. Moreover, if changes are not communicated to the users, it might cause complications in their use of open data, where consequences can be hard to anticipate. Based on the identified literature, we see that evaluation and feedback need research, especially in relation to barriers and on both active and passive evaluation.

5 Conclusions and Future Work

In this paper, we have identified and systematised OGD barriers into identifying data suitability, release decisions, data publishing, data usage, and evaluation based on a systematic literature review. We have illustrated the attention given to different barriers in OGD research from a longitudinal perspective and illustrated different types of barriers and focuses in the research field, including relating 46 OGD barriers to different parts and phases of the OGD process. With the illustrations in Sect. 3, have we positioned and contextualised our research questions, and with the systematisation (Sect. 4), we have integrated the research questions to contribute to a result that is valuable for researchers in the area, and in the long run for the practice. In the analysis, we moved away from a perspective on barriers as something to be "easily" reduced or

managed (e.g. regarding deciding to release OGD), towards viewing barriers as a network in a cycle and part of a strategic decision in a broader context. This is a way to expand the scope of OGD, where we move away from the reported myths surrounding open data.

We have also identified a knowledge gap between use and evaluation for barriers. The government can not solve barriers here, only mitigate by, e.g. ease usage. Our study opens for research from both the individual perspective and systematic perspectives on OGD barriers and their relations. One important and unexplored dimension of this; is the responsibilities of and division of labour between the publisher and user. We have also identified a need for further insight into open data barriers and myths, especially on the economic aspects, assumptions, and the gap between use and evaluation.

This paper indicates two research implications. (1) The reviews [e.g. 3, 17, 42] are essential to identify and understand OGD barriers. To hold the knowledge body together, we need recurrent systematic literature reviews both on a general and specific level (e.g. usage or publication), where authors offer insights, research avenues, and address research barriers. One such avenue for a future systematic literature review is to investigate the focuses in previous OGD barrier studies. Hopefully, this will open for successful research and practice. (2) Current research does not support informed decisions towards a certain kind of usage; current research is rather general or even abstract. We need a thorough understanding of what barriers a publisher will encounter, but we lack the how and why. There is time to move to answer those questions and combine different conclusions to support and guide practice.

We have identified some limitations in our study. We did not use "open government data" as a keyword of the search. Instead, we used the broader "open data". The broader term usage results in that not all the literature were studies of governmental nature but may be relevant for that sub-field. We also used singulars instead of plural, because of this we did not identify [8]. We looked at the top 600 hits from Google Scholar, filtered with the year and different keywords. To avoid being too limited, we used snowballing, but there might still be important literature out there. Another limitation is that the systematisation is not yet evaluated in practice. We also acknowledge that barrier severity varies [e.g. 5]. As the systematisation is based on a broad pool of barrier research, we claim that the coverage, and to a degree the usability, of the systematisation is high, but taking the limitations into account more work can be done to evaluate and validate it.

References

1. Albano, C.S., Reinhard, N.: Open government data: facilitating and motivating factors for coping with potential barriers in the Brazilian context. In: Janssen, M., Scholl, H.J., Wimmer, Maria A., Bannister, F. (eds.) EGOV 2014. LNCS, vol. 8653, pp. 181–193. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44426-9_15
2. Appelgren, E., Nygren, G.: Data journalism in sweden: introducing new methods and genres of journalism into "old" organizations. Digit. Journal. **2**(3), 394–405 (2014)
3. Attard, J., Orlandi, F., Scerri, S., Auer, S.: A systematic review of open government data initiatives. Gov. Inf. Q. **32**(4), 399–418 (2015)

4. Barry, E., Bannister, F.: Barriers to open data release: a view from the top. Inf. Polity **19**(1–2), 129–152 (2014)
5. Beno, M., Figl, K., Umbrich, J., Polleres, A.: Open data hopes and fears: Determining the barriers of open data. In: 2017 Conference for E-Democracy and Open Government (CeDEM), pp. 69–81 (2017)
6. Brugger, J., Fraefel, M., Riedl, R., Fehr, H., Schöneck, D., Weissbrod, C.S.: Current barriers to open government data use and visualization by political intermediaries. In: Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016, pp. 219–229 (2016)
7. Charalabidis, Y., Alexopoulos, C., Loukis, E.: A taxonomy of open government data research areas and topics. J. Organ. Comput. Electron. Commer. **26**(1–2), 41–63 (2016)
8. Conradie, P., Choenni, S.: Exploring process barriers to release public sector information in local government, p. 5 (2012)
9. Conradie, P., Choenni, S.: On the barriers for local government releasing open data. Gov. Inf. Q. **31**(SUPPL.1), S10–S17 (2014)
10. Cronin, P., Ryan, F., Coughlan, M.: Undertaking a literature review: a step-by-step approach. Br. J. Nurs. **17**(1), 38–43 (2008)
11. Davies, T.: Open data, democracy and public sector. Interface, pp. 1–47 (2010)
12. Dawes, S.S., Vidiasova, L., Parkhimovich, O.: Planning and designing open government data programs: an ecosystem approach. Gov. Inf. Q. **33**(1), 15–27 (2016)
13. Hellberg, A.S., Hedström, K.: The story of the sixth myth of open data and open government. Transform. Gov.: People, Process Policy **9**(1), 35–51 (2015)
14. Hjalmarsson, A., Johannesson, P., Juell-Skielse, G., Rudmark, D.: Beyond innovation contests: a framework of barriers to open innovation of digital services. In: Twenty Second European Conference on Information Systems, pp. 1–15 (2014)
15. Hjalmarsson, A., Johansson, N., Rudmark, D.: Mind the gap: exploring stakeholders' value with open data assessment, vol. 2015, pp. 1314–1323, March 2015
16. Hogan, M., et al.: Governance, transparency and the collaborative design of open data collaboration platforms: understanding barriers, options, and needs. In: Ojo, A., Millard, J. (eds.) Government 3.0 – Next Generation Government Technology Infrastructure and Services. PAIT, vol. 32, pp. 299–332. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63743-3_12
17. Hossain, M.A., Dwivedi, Y.K., Rana, N.P.: State-of-the-art in open data research: insights from existing literature and a research agenda. J. Organ. Comput. Electron. Commer. **26**(1–2), 14–40 (2016)
18. Huang, R., Lai, T., Zhou, L.: Proposing a framework of barriers to opening government data in China. Libr. Hi Tech **35**(3), 421–438 (2017)
19. Jaakkola, H., Mäkinen, T., Eteläaho, A.: Open data: opportunities and challenges. In: Proceedings of the 15th International Conference on Computer Systems and Technologies, CompSysTech 2014, pp. 25–39 (2014)
20. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. **29**(4), 258–268 (2012)
21. Kassen, M.: A promising phenomenon of open data: a case study of the Chicago open data project. Gov. Inf. Q. **30**(4), 508–513 (2013)
22. Kassen, M.: Open data in Kazakhstan: incentives, implementation and challenges. Inf. Technol. People **30**(2), 301–323 (2017)
23. Kucera, J., Chlapek, D.: Benefits and risks of open government data. J. Syst. Integr. **5**, 30–41 (2014)
24. Lindman, J., Kinnari, T., Rossi, M.: Business roles in the emerging open-data ecosystem. IEEE Softw. **33**(5), 54–59 (2016)

25. Lundell, B.: e-governance in public sector ict procurement: what is shaping practice in Sweden? Eur. J. ePractice **12**, 66–78 (2011)
26. Martin, C.: Barriers to the open government data agenda: Taking a multi-level perspective. Policy Internet **6**(3), 217–240 (2014)
27. Martin, S., Foulonneau, M., Turki, S., Ihadjadene, M.: Open Data: Barriers, Risks and Opportunities, pp. 301–309 (2013)
28. Melin, U.: Challenges and benefits in an open data initiative - a local government case study of myths and realities. In: 15th IFIP Electronic Government and the 8th Electronic Participation Conference (EGOV ePart 2016), vol. 23, pp. 111–122 (2016)
29. Melin, U., Lindgren, I.: Open government data in an e-service context - managerial and conceptual challenges. In: Americas Conference on Information Systems, vol. 23, pp. 1–10 (2017)
30. Rohunen, A., Markkula, J., Heikkila, M., Heikkila, J.: Open traffic data for future service innovation - addressing the privacy challenges of driving data. J. Theor. Appl. Electron. Commer. Res. **9**(3), 71–89 (2014)
31. de Rosnay, M.D., Janssen, K.: Legal and institutional challenges for opening data across public sectors: towards common policy solutions. J. Theor. Appl. Electron. Commer. Res. **9** (3), 1–14 (2014)
32. Saldaña, J.: The coding manual for qualitative researchers. Sage, Thousand Oaks (2015)
33. Susha, I., Grönlund, A., Janssen, M.: Driving factors of service in- novation using open government data: an exploratory study of entrepreneurs in two countries. Inf. Polity **20**(1), 19–34 (2015)
34. Susha, I., Grönlund, K., Janssen, M.: Organizational measures to stimulate user engagement with open data. Trans. Gov.: People, Process Policy **9**(2), 181–206 (2015)
35. Susha, I., Johannesson, P., Juell-skielse, G.: Open data research in the nordic region: towards a scandinavian approach? Int. Fed. Inf. Process. **9820**, 61–73 (2016)
36. Walters, W.H.: Comparative recall and precision of simple and expert searches in google scholar and eight other databases. portal: Libr. Acad. **11**(4), 971–1006 (2011)
37. Wirtz, B.W., Piehler, R., Thom-as, M.J., Daiser, P.: Resistance of public personnel to open government: a cognitive theory view of implementation barriers towards open government data. Public Manag. Rev. **18**(9), 1335–1364 (2016)
38. Yang, T.M., Lo, J., Shiang, J.: To open or not to open? determinants of open government data. J. Inf. Sci. **41**(5), 596–612 (2015)
39. 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. PAIT, vol. 4, pp. 115–135. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-9563-5_8
40. Zuiderwijk, A., Janssen, M.: The Negative Effects of Open Government Data - Investigating the Dark Side of Open Data, pp. 147–152 (2014)
41. Zuiderwijk, A., Janssen, M.: Towards decision support for disclosing data: closed or open data? Inf. Polity **20**(2–3), 103–117 (2015)
42. Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.S.: Socio-technical impediments of open data. Electron. J. e-Government **10**(2), 156–172 (2012)
43. Andrade, C.: How to write a good abstract for a scientific paper or conference presentation. Indian J. Psychiatry **53**(2), 172 (2011)
44. Jacinto, T., et al.: Doing science: writing conference abstracts. Breathe **10**(3), 265–269 (2014)

Open Government Data Driven Co-creation: Moving Towards Citizen-Government Collaboration

Keegan McBride(✉), Maarja Toots, Tarmo Kalvet, and Robert Krimmer

Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Tallinn, Estonia
Keegan.mcbride@ttu.ee

Abstract. It is believed that co-creation may lead to public service quality improvements, the provision and creation of new and innovative services, and bring public service providers closer to their service users. There has been an increased interest and focus on how new technological innovations are enabling and facilitating co-creation; one such digital innovation is open government data (OGD). This paper examines a relatively new concept, that of co-created OGD-driven public services and aims to understand how the availability and exploitation of OGD to co-create new public services allows service users to become collaborators rather than customers of public service providers. An exploratory case study is conducted on a pilot project within Estonia where a new public service has been co-created through the exploitation of OGD. The initial results show that in order for an OGD-driven public service to be effectively co-created, a new understanding of the role of stakeholders is needed.

Keywords: Open government data · Co-creation · Public service innovation

1 Introduction

Currently, within the fields of e-Government and public administration, scholars are proclaiming that public services are changing, Information Communication Technologies (ICTs) are helping to drive this change, and ICTs are empowering citizens, which may lead to higher rates of innovation in the public sector [1–5]. Similarly, there is talk of changing how we understand and define a public service; for the purpose of this paper, the following definition of public service is adopted: public services are "any services which are offered to the general public with the purpose of developing public value, regardless of the role that the public sector plays in the process" [6].

One major way that ICTs are empowering citizens and transforming public service creation and delivery is through digitally enabled co-creation. Specifically, it seems that ICTs and digital technologies can be applied in a variety of ways to enable different forms of co-creation [7]. A recent paper notes that digital technologies and ICTs are primarily effecting co-creation in three different ways: changing traditional co-creation, enabling new forms of co-creation, and replacing traditional co-creation through

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 184–195, 2018.
https://doi.org/10.1007/978-3-319-98690-6_16

automated processes [7]. However, as noted by [8], more work is needed when it comes to understanding how OGD may be used to create new and innovative services; it is also brought up by [9] that there is little empirical work and more research is needed into the emerging phenomenon of co-created OGD-driven public services. A recent paper, [10], discusses the relationship between OGD and co-creation, but the concept of co-created OGD-driven public services is not touched on in detail. To better address the research gap, two research questions were formulated that this paper aims to address: (1) How can OGD contribute to the co-creation of new public services? (2) How does OGD influence our understanding of stakeholder roles in the public service delivery process?

In order to answer these research questions, an exploratory case study was conducted on a pilot project that is being conducted as part of a Horizon2020 (H2020) project where a new public service is being co-created through the exploitation of OGD. This project aims to explore how OGD may be used to help drive co-creation and innovation within the public sector [11]. To demonstrate how this happens, six pilots are being conducted, one of which is taking place in Estonia. The Estonian Real Estate Pilot Program (EREPP) is a pilot project that is carried out in cooperation between the Estonian Ministry of Economic Affairs and Communications (MKM) and Tallinn University of Technology (TTU). The authors of this paper were directly involved with the project and in charge of the implementation of the pilot application in Estonia. This project gives the researchers direct access to the process of co-creating a new public service that is based on OGD; for this reason the project was selected as an appropriate case for answering the proposed research questions.

The paper starts with an initial overview of the concepts of co-creation, OGD, and OGD-driven co-created public services. Following this, the methodology for the paper is put forth and, additionally, a conceptual model is provided as it aids the investigation of co-created OGD-driven public services. In the next section, the context surrounding the case and the case itself is presented. Following this, a discussion of the results of the case will take place where propositions will be forth in addition to reflections and implications that this case may have for the current theory and understanding of co-created OGD-driven public services.

2 Co-creation and OGD

Co-creation. The term co-creation has its roots in the concept of "coproduction", which was first coined by Elinor Ostrom 1972 [12]. Ostrom found that in areas where citizens were more forthcoming and welcoming to law enforcement, there was a higher level of public service, or a higher production of public value, compared to areas where citizens were not as cooperative with the police [13, 14]. She thus concluded that the value of a public service was very much determined by not just the provider of the service, but by the interaction between the consumer of the service and the provider [13, 15, 16]. When Ostrom talks about coproduction, she defines it as "the process through which inputs used to produce a good or service are contributed by individuals who are not "in" the same organization" [17]. Ostrom also notes that using the term "client" when defining a service is not necessarily the best term as client is "a passive

term", and in her understanding of coproduction citizens can "play an active role in producing public goods and services of consequences to them" [17].

Though participation of service consumers is paramount for the success of a public service, one should not count on service consumers to be automatically engaged and active citizens once a new service is provided [18]. A new public service needs to motivate active coproduction; however, if a public service requires higher levels of motivation for participation it is also likely that there will be an increased effort "required of service consumers to overcome hurdles to participation" [19]. One way to lower the barriers to coproduction is to involve citizens at every stage of the public service creation process; this is known as co-creation. In this paper, the term "co-creation" may be understood as "the involvement of outside, non-typical, stakeholders in the initiation, design, implementation, and evaluation of a new public service" [6].

Open Government Data. This paper deals specifically with OGD rather than open data in general. There are currently many different definitions of OGD [8, 20, 21], but most definitions share some core components: data must be machine readable, it should be licensed in a way to allow easy sharing and reuse of data, and it should be usable and understandable by humans. With this in mind, for the purpose of this paper the following definition of OGD will be used: OGD is non-confidential data which is gathered, and subsequently released by a government organization in a machine readable format which is discoverable, usable, and freely available [8, 21, 22].

Benefits of OGD. There have been recent academic works such as [8, 22] that present some of benefits that may be provided by OGD. Some benefits (though there are likely many more) are increased transparency, new forms of social participation, innovation, creation of new public services, increased accountability, creation of new business models and improved data models [8], [23–25].

Barriers Relating to the Use and Release of OGD. If OGD is released, and it is truly open (it meets the requirements set out in the definition provided earlier), then it has the potential to create major benefits for society. However, as pointed out by [8], just making OGD available is not enough, as "the value is created by its use". OGD usage generally refers to any interaction an actor (a user of OGD can come from any sector be it private or governmental [22]) has with the data, such as downloading, analyzing, or exploiting the data [9, 26, 27].

There have been many attempts to provide a better understanding of OGD barriers [8, 22, 28–30]. On the user level, commonly cited barriers are the lack of technological understanding/ability; lack of domain knowledge; language barriers to the data; lack of time to use data [22, 28, 30, 31]. On the government level, commonly cited barriers are missing political motivation; no understanding of the potential benefits of open data; missing technical infrastructure or technical know-how; poor data quality; confidentiality or personal data issues related to the release of data [8, 22, 28, 30, 31].

On the government level, a majority of the barriers are directly related to the release and publishing of open data whereas, on the user level, a majority of the barriers relate to their ability to use or understand OGD. However, in the case where these user barriers do not apply, namely when the user of the data has a strong understanding of data analytics and a personal interest in open data, they may often struggle to use the

OGD which is provided. In this situation, the most relevant barriers are related to the poor quality (encoding issues, missing values, lack of metadata, etc.) of the OGD, lack of interesting information, outdated data, and lack of an application programming interface (API) functionality [31].

3 Framework for Understanding OGD-Driven Co-created Public Services

In order to better understand this concept of co-created OGD-driven public services, the framework (Fig. 1) put forth by [6] is to be used.

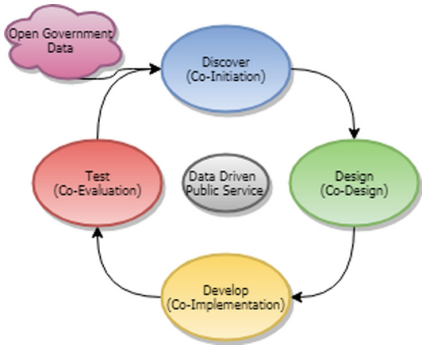

Fig. 1. OGD-driven Public Service Co-Creation Framework. Source: [6]

In previous research, the authors of the framework proposed that OGD plays an almost catalytic role in the co-creation process and, once OGD has been introduced, it may be exploited by any actor to co-create a new public service. This framework draws on the ideas of agile development, lean methodology, and co-creation to propose an OGD-driven co-creation cycle, which consists of four steps: co-initiation, co-design, co-implementation, and co-evaluation. This conceptualization of the process puts forth the idea that when OGD is released, any actor is able to exploit OGD and become an active participant in the co-creation of a new public service. This process is able to start when access to OGD becomes available and OGD becomes subsequently exploited.

In order to understand better how OGD exploitation occurs, four propositions have been put forth which will then be referred back to in the discussion.

- Proposition 1 (P1): In order for OGD to be able to contribute effectively to the co-creation of new public services, a change in the current understanding of stakeholders in the public service creation process must take place.
- Proposition 2 (P2): If OGD is released by government agencies, and this data is used to create a new public service, then the government is a participant in its co-creation.

- Proposition 3 (P3): OGD may be used as a base, or platform, from which new and innovative public services may be co-created between government agencies, who maintain and release the data, and with outside stakeholders.
- Proposition 4 (P4): OGD has the potential to enable, encourage, and drive public service co-creation.

4 Methodology

The research for this paper was carried out in the form of a single holistic exploratory case study [32]. For this paper, four different sources of evidence have been used: semi-structured interviews, workshops, survey, and documents. Six semi-structured interviews were conducted with members from Estonian public administration and from Estonian civil society; all members were selected due to the interest and role in OGD and service design and delivery. A list of five basic questions was drafted and used as the starting point for all interviews and all interviews were recorded, transcribed, coded, and then analyzed. Five interviews were done in person, and one interview was done through email. Two different workshops were attended (the first of which was organized by the authors). The workshop organized by the authors followed Nominal Group Technique [37]; in attendance were seven different public sector agencies and two Estonian start-ups; all participants played a role in either releasing, maintaining, or using OGD. The second workshop was a working group made up of members from private sector companies and public sector organizations in Estonia that were working together to plan and understand how government could become more data-driven. In the workshops, all responses to questions were written down by the participants and collected at the end of the day by the organizers, and then coded and analyzed. As part of a larger study, a survey was conducted that aimed to gain insight into how OGD could be used in the co-creation on new public services. This survey received 63 responses, however the responses received represented six different countries so only the responses from the Estonian stakeholders were included in this paper. Nine responses were received from Estonian stakeholders: three from the non-governmental sector and six from the public administration sector; for previous discussion on the results of the survey, please see [22]. Finally, for the document analysis, Estonian public policy documents, white papers, and laws were examined. For the purpose of anonymity, interview and survey responses are cited in this paper using codes: interviews are referred to by letters from "A" to "F" and survey respondents by four-letter codes where "EE" indicates Estonia, and "PA" or"NG" indicate whether the respondent represents the public sector or the non-governmental sector respectively.

5 The Case

5.1 Case Context

The pilot case was part of the H2020-funded project OpenGovIntelligence which had selected six different countries for pilot projects to be carried out. These pilot projects

should address a relevant and current societal issue and, if done successfully, the pilot should demonstrate how OGD is able to drive innovation, improve service quality, and overcome barriers relates to the use and exploitation of OGD in the public service creation context [11]. For the project, Estonia was able to provide a problem, potential datasets, and articulate a reasonable solution for the problem. Additionally, Estonia also came with the image of being a world leader and expert when it comes to ICT innovation, e-services, and e-society [33]. Thus, Estonia was expected to provide a good location for using OGD, in combination with newly created ICT tools, to create a new public service that has a high potential to create new benefit for society.

Once Estonia had been selected as a pilot country, the process to understand the current situation of OGD and co-creation began. The initial research included an overview of the potentially relevant legal texts, an overview of relevant government ministry documents, a survey, and further on in the project unstructured interviews were undertaken to get further knowledge of stakeholder perception of OGD and co-creation of publics services. The results of this research provide the necessary contextual foundation for understanding a more in depth discussion of the case. Inside Estonia OGD is regulated by the Public Information Act (PIA). In the act it is stated that all data that may be used for public purposes, that is to say it is not restricted by law, shall be opened to the public [34]. The PIA continues on to say that data should be released in machine readable format and come without any restriction on reuse of the data; however, it also states that this is only required if it would not involve "disproportionately great effort" for data holders [34]. Agencies should be following an "open by default" policy, but often what ends up happening is agencies hire a person to deal with all incoming data requests (Interview D). Estonia has made steps in the right direction for OGD, but as it is not a concrete requirement for all data to be made available in an open and machine-readable format, the availability of OGD in Estonia is lacking (Interviews A&C, Survey EENG1, EENG3, EEPA5).

The second important part of the background information relates to the current situation and understanding of co-creation of public services within Estonia. The idea of co-creation of public services within Estonia is one that does not have much government support (Interview A-F, Survey EENG1, EENG2, EENG3, EEPA3, EEPA5, EEPA6). Some of the reasons for this include not enough funding, not enough citizen demand, low levels of collaboration between citizens and government, and lack of understanding of the concept (Interviews A, B, C, D, E, F). With this in mind, steps are being taken to try to move towards creation of new public services that have been co-created with citizens (Interview E & F). Many agencies are going out to end users to ask for their input on what services they need and then trying to involve them in the design and creation process (Interview D).

Though Estonian public agencies are beginning to consult potential service end users at the beginning stages of development, this is often where the cooperation ends (Interview B). When looking at the usage of OGD for the co-creation of new public services inside Estonia there is not, yet, an example as it is not currently occurring. Some of the main reasons for this are the infancy stage of the Estonian OGD infrastructure and agencies not being willing to participate in co-creation of services with citizens and other stakeholders (Interviews B & F).

5.2 Case Description

Due to the poor quality of OGD in Estonia, one of the aims of the pilot project was to lobby for increased availability of OGD. The pilot project aimed to include users in the service design, development, implementation, and evaluation as often as possible. The events presented within this case study took place between June 2016 and December 2017 and represent the co-initiation, co-design, and co-implementation stage.

Co-initiation. The idea for the pilot project within Estonia was co-initiated. The initial topic was proposed by foreign academics and students within Estonia, the idea was then presented to the Ministry of Economic Affairs and Communications (MKM) where it was refined and accepted. The need for such a service has also been highlighted by numerous individuals in Estonian expat groups both in person and on social networks. Once the service had been accepted and put forth, the next steps were the co-design and co-implementation of the service.

Co-design. As the service aims to demonstrate how OGD may be used to create a new public service, the first task to be completed was to gain an overview of the OGD sets that were currently available. In order to identify these datasets, two different approaches were used. The first and initial starting point was a simple Google search using the Estonian term for open data, "Avaandmed". The second approach was to make requests to government agencies for datasets that could be useful for pilot development.

To aid in the initial design of the new service, two workshops were conducted. The first Estonian Real Estate Pilot Program Workshop was conducted on 16 September 2016. The workshop had nine attendees (excluding the organization team) who represented seven different government agencies and two private sector companies. The workshop was divided into two sessions. In the first session, the participants discussed and came up with four main benefits of an Estonian real estate portal based around OGD: fairer pricing, happier citizens, one stop shop for real estate data, and increased availability and usage of real estate information. The second session was titled "Constructing the Functionality" and dealt primarily with constructing user stories and personas. In any agile development project, personas and user stories play a critical role as they allow the development to reflect better the actual needs of the users [35]. Two core target groups consistent throughout the participants' work: foreign students and foreign employees who are moving to Tallinn were selected as the initial groups for the pilot.

Taking into account the personas and the user stories, the initial datasets that should be included in the pilot program were discussed, voted on, and selected. These five datasets were public transport data, safety data, price data, point of interest data (schools or doctor's offices), and property information (such as age of the building, amenities within the building, or the accessibility of the building).

A second presentation and workshop took place at a meeting of the Estonian Data Analytics Working Group, which is made up of members from multiple public and private sector organizations. What was discovered at the workshop, and matched the literature and the results of the previous workshop, was that OGD was generally looked upon favorably and as a needed innovation, but that there was no political will or user

demand for better OGD. The private sector representatives took a critical approach towards the presented service, the general criticism was that this work seemed to be better suited for private sector agencies and that the data that had thus far been cleaned would be very valuable for some private sector companies' business models.

Co-implementation After the workshops, the five OGD sets that had been selected for use were explored. These datasets required a large amount of cleaning and manipulation in order to make them usable, this initial work was conducted by staff members at TTU. At the same time this data cleaning was progressing, a hackathon sponsored by MKM that dealt with big and open data was announced. In order to test, design, and implement a new OGD driven service, the datasets that had been gathered and cleaned were brought to the hackathon (which took place from 21 to 23 October, 2016) and the idea for an OGD-driven real estate portal was pitched and selected.

A team was formed with members from the project team at TTU and members of the big data science team from the private sector company Nortal. As a compromise between these two groups, the datasets that had been cleaned and obtained by TTU's project team would be used, but it was to be a commercially oriented service rather than a free service; however, the creation of public value was still to be the main goal. Over the next 48 h, a new service MVP was built which used OGD to rate different addresses based on an individual's preferences. After 48 h, this idea was presented to the audience where it received an honorary mention for providing valuable location based information.

After the hackathon, development continued on the pilot project. Initially, MKM was intended to develop the new service in cooperation with TTU and outside stakeholders. However, there was much organizational push back, which eventually led to TTU taking the lead in pilot development. The initial development sprint took place between 15 and 18 March, 2017; the initial goal of the sprint was to develop a fairly simple and easy to understand user interface for the pilot project. As EREPP aims to encourage others to participate in the design and development of the service, the code is completely open source and hosted on GitHub.

6 Discussion

Earlier, four different propositions were put forth and investigated, the case will now be discussed, and the propositions reflected back upon. P1 stated that a change in the current understanding of stakeholders in the public service creation process must take place, if OGD was to be able to contribute effectively to the co-creation of new public services. Along with a change in understanding of the role of stakeholders, there must also be an organizational change in how public services are understood.

There are many different understandings of the definition of public service within Estonia, but one that is often referred to is as follows, "a public service is a service that the state, local government, or a person in private law performing public duties provides at the will (including the presumable will) of a person for the performance of their legal obligations or exercise of their rights" [36]. In the interviews it was claimed that a public service was, in essence, something that was paid for by public money and

carried out by a public agency (Interviews B,C,D,E,F). Thus, according to these definitions, citizens could not (should not) be able to create new public services. This understanding of public services also carries out into the understanding of the roles of the stakeholders in the public service creation process. When asked, interviewees often said that citizens should be consulted at the beginning stages of a new service and asked for feedback throughout (Interviews B,D,E,F). However, when asked if citizens should be able to play a role in the creation and design of a new public service, the answer appears to be no (Interviews B,F). The interviewees did state that though outside stakeholders are currently not able to play a role in public service co-creation, this may change in the future (Interviews B,F). There seem to be two primary reasons for why stakeholders are not currently viewed as being able to play a role in the co-creation of new public services in Estonia. The first relates back to the definition of a public service in Estonia, and the second is that citizens are referred to as clients or customers rather than as partners or collaborators (Interviews D,E,F). Throughout the case, resistance to the notion that government agencies could work with citizens as partners was clear. The clearest example of this is through the actions of MKM. During the case, TTU took over the pilot implementation from MKM. This was an interesting development, as TTU and MKM still worked together, but instead of the government agency developing the new public service, a university had taken the lead. Though a new public service is in the process of being created, and it does exploit some OGD sources, the effectiveness of the pilot program has been hindered by the lack of access to OGD and the organizational belief that outside stakeholders should not be able to play a part in the co-creation of new public services.

In regards to P2, there seems to be an interesting paradox currently in place in Estonia. On the one hand, some interviewees stated that citizens should not, or are not, able to play a role in the co-creation of a new public service (Interview B). On the other hand, government agencies do make some of their data open, and this data can be exploited to create public value (as demonstrated by this case). Thus, by releasing open data, government agencies are willingly becoming a participant in the co-creation of new services (whether they mean to or not). When government agencies release open data, citizens have the possibility to use, analyze, and exploit this data. In the Estonian case, government agencies were constantly communicated with to discuss issues in relation to data availability, data quality, data structure, etc. This communication accomplished a few different things: it increased communication between service developers and government agencies, it increased awareness of data issues, and, as noted by Interviewee A, these conversations help government agencies become aware of what data they have, what they do not, and what the current issues are.

Related to P3, by opening up datasets, government agencies allow other stakeholders to create public value through the exploitation of their data, while at the same time gaining valuable information in regards to their own data; thus, the government acts platform-like. Though this interaction takes place and does seem to provide tangible benefits for both government sector and other stakeholders, it is also one of the largest barriers present. The reason for this is that releasing data requires government agencies to acknowledge that there is an alternative way to create public services and that other stakeholders may come to be seen as partners or as collaborators rather than as customers or dependents. In line with P4, the case does seem to confirm that there is

a relationship between OGD and co-creation. If OGD is available, any actor is able to exploit or analyze this data to create new public value. Throughout the process where OGD becomes exploited, co-creation is occurring, at a minimum, between the government agency and the actor that is exploiting the data. Furthermore, the service provider is also acting a service user at the same time, as they are reliant upon the government's open data. Finally, if an application is developed on top of exploited OGD, a complex public service delivery system begins to emerge. These public service delivery systems are made up of many different stakeholders with different goals, motivations, and behaviors; based on this one case, it does seem to be true that networks, relationships, and feedback have a strong role in influencing the design, development, and implementation of the new public service.

7 Conclusion and Future Research

The aim of this paper was not to provide a thorough discussion and overview of all the drivers and barriers of OGD, but, rather, to explore and understand the process in which new public services can be co-created through the use and exploitation of OGD. The provision and creation of new and innovative services is one proposed benefit of OGD. Specifically, OGD seems to have the potential to play a catalytic role in the co-creation of new public services. Though this potential appears to exist, there is a large research gap and lack of empirical studies that aim to understand how this process actually occurs. To address this research gap, an exploratory case study was conducted to gain some initial empirical understanding of the process that is undergone to co-create an OGD-public service. An initial conceptual model for understanding co-created OGD-driven public services served as the foundational point for understanding this process.

As a result of this exploration, some interesting conclusions seem to appear. Firstly, the availability of OGD may lead to a change in our understanding of public service delivery and the roles that different stakeholders play in this process. In order for the co-creation of OGD-driven public services to take place, governments must be willing to acknowledge that non-traditional stakeholders can take the role of public service creator; if this does not occur, then it becomes increasingly difficult for other stakeholders to co-create new and innovative services. Secondly, it appears that there is genuine interest from non-traditional stakeholders to get involved in the co-creation of new public services; they just need to be given the opportunity to do so. This was shown by the participation of many stakeholder groups throughout the design of the Estonian pilot. Finally, some barriers that may inhibit the co-creation of OGD-driven public services that have been brought out in the literature seem to be reaffirmed: low data quality, organizational push back, inadequate legal frameworks, and a lack of government support.

Though the findings that have emerged from this paper seem to be confirmed from this case study, it must be noted that this is a single case study and, as such, further empirical research should be conducted to test these conclusions and recommendations. This future research could aim to understand how OGD influences public service co-creation, the sustainability of co-created OGD-driven public services, how the process differs depending on what stakeholder group drives the process, and how the presence

of OGD influences the relationships between public service producers and public service consumers.

Acknowledgements. This work was supported by the European Commission (OpenGovIntelligence H2020 grant 693849), Estonian Research Council (PUT773, PUT1361) and TTU Digital Governance Competency Center (SS483).

References

1. European Commission: A vision for public services (2013)
2. OECD: Together for Better Public Services: Partnering with Citizens and Civil Society. OECD Publishing (2011)
3. Dunleavy, P., Margetts, H., Bastow, S., Tinkler, J.: New public management is dead - Long live digital-era governance. J. Public Adm. Res. Theory **16**, 467–494 (2006)
4. Linders, D.: From e-government to we-government: defining a typology for citizen coproduction in the age of social media. Gov. Inf. Q. **29**, 446–454 (2012)
5. Osborne, S.P., Radnor, Z., Nasi, G.: A new theory for public service management? toward a (public) service-dominant approach. Am. Rev. Public Adm. **43**, 135–158 (2013)
6. Toots, M., et al.: A framework for data-driven public service co-production. In: Janssen M., et al. (eds.) Electronic Government, EGOV-EPART 2017, St.Petersburg, pp. 1–13 (2017)
7. Lember, V.: The increasing role of new technologies in co-production. In: Brandsen, T., Steen, T., Verschuere, B. (eds.) Co-production and Co-creation: Engaging Citizens in Public Service Delivery. Routledge, Abingdon (2018)
8. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. **29**, 258–268 (2012)
9. McBride, K.D., Aavik, G., Kalvet, T., Krimmer, R.: Co-creating an open government data driven public service: the case of Chicago's food inspection forecasting model. In: 51st Hawaii International Conference on System Sciences (HICSS), pp. 2453–2462. IEEE (2018)
10. Khayyat, M., Bannister, F.: Towards a model for facilitating and enabling co-creation using open government data. Inf. Polity **22**, 211–231 (2017)
11. Krimmer, R., Kalvet, T., Toots, M., McBride, K.: Deliverable 2.1 OpenGovIntelligence framework–first release (2016)
12. Ostrom, E.: Metropolitan reform: Propositions derived from two traditions. Soc. Sci. Q. **53**, 474–493 (1972)
13. Ostrom, E., Parks, R.B., Whitaker, G.P., Percy, S.L.: Formation of police and law enforcement policy. Policy Stud. J. **7**, 381–389 (1978)
14. Osborne, S.P., Radnor, Z., Strokosch, K.: Co-production and the co-creation of value in public services: a suitable case for treatment? Public Manag. Rev. **18**, 639–653 (2016)
15. Sharp, E.B.: Toward a new understanding of urban services and citizen participation: the coproduction concept. Am. Rev. Public Adm. **14**, 105–118 (1980)
16. Whitaker, G.P.: Coproduction: delivery in service participation citizen. Public Adm. Rev. **40**, 240–246 (1980)
17. Ostrom, E.: Crossing the great divide: synergy, and development. World Dev. **24**, 1073–1087 (1996)
18. Pestoff, V.: Co-production as a social and governance innovation in public services. Polityka Spoleczna **11**, 2–8 (2015)
19. Verschuere, B., Brandsen, T., Pestoff, V.: Co-production: the state of the art in research and the future agenda. Voluntas **23**, 1083–1101 (2012)

20. Gonzalez-Zapata, F., Heeks, R.: The multiple meanings of open government data: understanding different stakeholders and their perspectives. Gov. Inf. Q. **32**, 441–452 (2015)
21. Open Knowledge Foundation: What is Open Data?. http://opendatahandbook.org/guide/en/what-is-open-data/
22. Toots, M., McBride, K., Kalvet, T., Krimmer, R.: Open data as enabler of public service co-creation : exploring the drivers and barriers. In: International Conference for E-Democracy and Open Government, Krems, Austria, pp. 102–112. IEEE (2017)
23. Janssen, M., Charalabadis, Y., Krcmar, H.: Open data, information processing and datification of government. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 2668–2670 (2017)
24. Dawes, S.S., et al.: Planning and designing open government data programs: an ecosystem approach. Gov. Inf. Q. **33**, 15–27 (2016)
25. Hartog, M., Mulder, B., Spée, B., Visser, E., Gribnau, A.: Open data within governmental organisations: effects, benefits and challenges of the implementation process. JeDEM - eJournal eDemocracy Open Gov. **6**, 49–61 (2014)
26. Foulonneau, M., Martin, S., Turki, S.: How open data are turned into services? In: Snene, M., Leonard, M. (eds.) IESS 2014. LNBIP, vol. 169, pp. 31–39. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04810-9_3
27. Foulonneau, M., Turki, S., Vidou, G., Martin, S.: Open data in service design. Electron. J. e-Government. **12**, 99–107 (2014)
28. Zuiderwijk, A., Janssen, M.: Open data policies, their implementation and impact: a framework for comparison. Gov. Inf. Q. **31**, 17–29 (2014)
29. Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.S.: Socio-technical impediments of open data. Electron. J. Electron. Gov. **10**, 156–172 (2012)
30. Conradie, P., Choenni, S.: On the barriers for local government releasing open data. Gov. Inf. Q. **31**, S10–S17 (2014)
31. Young, M., Yan, A.: Civic hackers' user experiences and expectations of seattle's open municipal data program. In: Proceedings of 50th Hawaii International Conference on Systems Science, pp. 2681–2690 (2017)
32. Yin, R.K.: Case Study Research: Design and Methods. Sage publications, Newbury Park (2013)
33. Kalvet, T.: Innovation: a factor explaining e-government success in Estonia. Electron. Gov. Int. J. **9**, 142–157 (2012)
34. Riigikogu: Public Information Act. Estonia (2000)
35. Chamberlain, S., Sharp, H., Maiden, N.: Towards a framework for integrating agile development and user-centred design. In: Abrahamsson, P., Marchesi, M., Succi, G. (eds.) XP 2006. LNCS, vol. 4044, pp. 143–153. Springer, Heidelberg (2006). https://doi.org/10.1007/11774129_15
36. Estonian Ministry of Economic Affairs and Communications: Green Paper on the Organisation of Public Services (2013)
37. Delbecq, A.L., de Ven, A.H., Gustafson, D.H.: Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes. Scott Foresman, Glenview (1975)

Exploring Open Data State-of-the-Art: A Review of the Social, Economic and Political Impacts

Shefali Virkar(✉) and Gabriela Viale Pereira

Danube University Krems, 3500 Krems, Austria
{shefali.virkar,
gabriela.viale-pereira}@donau-uni.ac.at

Abstract. This research paper explores the impact that the access to, and free usage of, stored (mainly public sector) data has on society, the economy and on good governance, together with the implications of this new paradigm for modern-day governments. Although a number of recent research studies attempt to identify the benefits and drawbacks of open data, or to demonstrate its role in governance processes, there exists to-date no systematic attempt to both rigorously survey current literature to enumerate the areas in which open data has had an impact and to discuss its potential as a significant tool for policymaking. To address this research gap, and to highlight its intrinsic value to different actors, we examine the current state-of-the-art on the impacts of open data research and practice through a systematic survey of extant scholarly and practitioner literature. The first part of our study will present a comprehensive overview of the societal, economic, and political impacts of open data. We will then evaluate the new possibilities afforded by open data to policymaking, and conclude with a discussion of its role in open governance.

Keywords: Open data · Open government data · Effects of open data

1 Introduction

As digital technologies come to be applied in different domains and sectors, most especially in the public sector, new possibilities for innovating governance have been created by the growth in data, computational power and social media [1]; all of which signal a new Era that is data-driven. In this scenario, open data emerges, being "one step further the knowledge-based society and economy" [2].

Open Government Data (OGD) can be considered as a specific subset of data which lies on the open data and government data domains [3]. Open Government has OGD as one of its principles, but requires also an extensive transformation of the public sector and its relationship with the public that needs to be mediated by information and communication technology (ICT) [4]. The principles of open government are data openness, transparency, participation, and collaboration [5]. Based on these principles, the idea of open government is to establish a modern cooperation among politicians, public administration, industry and private citizens [6].

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 196–207, 2018.
https://doi.org/10.1007/978-3-319-98690-6_17

In order to highlight the value of open data, the objective of this study is to present a comprehensive overview from the academic literature and practice on the impacts of open data along the following dimensions: (1) Social Impacts, (2) Economic impacts, and (3) Impact on Good Governance. The section on the creation of social capital covers the influence that the opening of data has on society and the rights of the individual. For instance, citizen empowerment and democracy are identified as core expected effects of OGD initiatives [7]. The second topic, economic impact, is often measured in monetary terms and the generation of revenue. For example, when analysing the effects of OGD utilization [8], economic development is considered the second most prominent aspect. Within the third domain, the role played by open data in the furtherance of good governance is discussed. Here transparency and accountability are listed amongst the primary effects of OGD initiatives [8].

The main theoretical contribution of this study pertains to the provision of an overview of the current state-of-the-art in open data research as one of the first comprehensive reviews to have a focus on the various impacts of OGD. The practical implications affect policymakers and governmental institutions by providing recommendations on how to maximize the benefits of OGD implementation, especially but not limited to any specific national context.

This paper starts by presenting a review of open data state-of-the-art and the methodology applied to this study. The research design is followed by the results and analysis of the literature review. The following section then presents policy recommendations. Finally, concluding remarks and an agenda for future research are discussed.

2 A Review of Open Data State-of-the-Art

Despite the clear potential of data for innovation [9, 10] and the growing amount of data that have been opened by several levels of government around the world, there are still doubts as to whether OGD will achieve all the expected benefits [11]. Open data has become a topic of increasing importance in research and practice, with a particular expansion in recent years [8, 12]. Despite being investigated with greater focus on public sector settings, open data studies are not exclusively related to government but also conducted in private settings [2]. "The foundation of the information provided by the government is referred to as open data, sets of data published by government that can be read and interpreted by either humans or machines" [13]. This study pays particular attention to data opened by governmental organizations, but we use the terms OGD and open data interchangeably for convenience.

The main impacts of open data initiatives in local governments are related to the use of data to improve decision-making and to better meet the needs of citizens (user side), as well as promoting transparency (provider side), due to the fact that governments provide access to and stimulate the re-use of public sector information [2, 14]. Transparency was also identified by [15] as one of the main objectives of opening government data, together with the direct impact provided by information access, increase in accountability and new possibilities for citizens to participate in the governance process.

Three recent literature reviews explore the OGD topic; focusing on the utilization of OGD [8], on assessing OGD initiatives [15] and providing insights from open data research [2]. Although the first two provide some discussion about the impact of OGD on stakeholders and main effects of OGD utilization, an overview of the current state-of-the-art in open data and its impact is lacking. The same applies to benchmarking studies such as the Global Open Data Index (GODI)[1] that focus on open government data publication from a civic perspective, but do not look at other aspects such as context, use or impact.

Proponents of OGD believe that the new role of the public sector as an information provider will have an impact on the creation of social and economic value and on the attainment of good governance [8]. Cultural challenges of open government data initiatives have strong relation to awareness about the benefits and potential of open data [15].

3 Methodology

In order to analyse the existing knowledge on the impacts of open government data we conducted a systematic literature review as recommended by [16], based on the framework proposed by [17]. According to [16], a systematic literature review involves "…identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest." In agreement with [16], we adopted a pre-defined search strategy to ensure that our literature review was thorough, fair, and of robust scientific value. Following [17], the design of the review for this paper consisted of: defining the search terms; selecting the sources; applying the search in the sources; going backward by reviewing the citations for the articles, selecting the sample based on the abstracts and finally based on the full-text.

The Scopus database (a comprehensive, reputable database for peer-reviewed scholarly and practitioner literature) and the Government Information Quarterly Journal (recognized as one of the main sources of literature in the government community) were selected as the main sources for the systematic literature review. Conventional key word searches using the search terms "open data" (6,407 articles) and "open government data" (481 articles) were conducted as the first step. Refinement of these results by publication date (only recently published articles from 2010 to the present), and subject relevance (proximity of abstract to topic based on author judgement) yielded 537 papers. The second round of refinement included looking for terms such as "impact", "effects", "implications" within the title, abstract, or keywords of this list for identifying specific aspects of open data. This step resulted in the identification of 80 papers. Finally, based on the citations from the identified articles in the first and second steps we determined further material that was also relevant for the analysis. The findings of the 60 most pertinent articles are summarised and discussed here.

[1] https://index.okfn.org/methodology/.

4 Findings on the Impacts of Open Data

One of the primary objectives of this research paper is to delineate and discuss in adequate detail the different domains within which open data bears significant effect so as to construct the foundation of a policy roadmap for the future release of information and knowledge. As discussed previously, a list of impacts pertaining to open data have been extracted through a systematic literature analysis of indexed scientific works, and are presented to the reader in this chapter. Our research findings are organized in the following manner: The first section is concerned with the social and political implications of open data. The second section deals with open data's economic implications. The third section outlines the contribution made by open data to good governance. Each derived proposition is then classified as one or the other, and listed, and then discussed.

4.1 Generating Social Value Through Open Data

Innovation in Public Services: Open data can foster innovation in public service provision [18], wherein the act of making information widely available increases the participation of a range of non-institutional actors who co-create [19] and renew public services [18] alongside the government. Considering that opening public data can result in improved public service standards, the provision of open data significantly changes the roles played by government [20] and citizens [21] in the production and consumption of these services. However, information availability and accessibility alone are not sufficient conditions to leverage social value from open data [19, 22]. Innovation in public services only arises through the re-use, dissemination and linking up of open government datasets [23]. The provision of tools and value-creation techniques must accompany any public release of data [19].

Fostering Citizen Empowerment: The release of open data contributes towards the empowerment of individuals within a society by altering the nature and scope of role that they play in the process of governance [24–28]. Citizen empowerment arises through a combination of access to open government data, and the availability of tools and technologies to combine it creatively. The provision of open data has altered the mindset of citizens towards government [24], actively engaging individuals in processes of governance [27]. Collaborative technologies and social software have further accelerated this change [28]. Thus, building an environment that encourages the emergence of networks of collaboration and co-creation, alongside the provision of open datasets, can translate into the empowerment of individuals within a society [26].

Promoting Meaningful Citizen Engagement: In providing the public with the data necessary to take well-informed decisions and actively engage with each other and with government, the release of open data has the potential to make citizen collaboration more meaningful [29–32]. Being able to regularly access high quality data about public services and government activities within an open data ecosystem will enable citizens to engage in meaningful dialogue with government [30, 32]. The open data ethos, through the free-access provision of information and tools, encourages proactive citizen

engagement by encouraging individuals to independently develop creative solutions to complex societal problems [32]. Government can guide these interactions, channeling efforts into priority areas, with active and regular public engagement in the co-creation of public services also being measured against the achievement of tangible societal and political outcomes [12].

Implications for Data Protection and Privacy: The opening up of government datasets for use by the general public has particularly significant ramifications for data protection and individual privacy [33–37]. While open data initiatives aim to create public value through (largely) unfettered innovation and transparent practices, increased openness may also lead to breaches in privacy and security violations as data potentially containing personally identifiable information (PII) is released into the public domain [33]. A balance needs to be struck between protecting the latter and leveraging the enormous potential benefit afforded by the public release of data [34, 36]. To prevent privacy breaches, it is necessary to eliminate all privacy-sensitive attributes prior to the opening of data [37].

4.2 Generating Economic Value Through Open Data

Encouraging Entrepreneurship: The relationship between open data and entrepreneurship is mutually reinforcing [38]; on the one hand, open data fosters entrepreneurship [39], whilst on the other, entrepreneurs create value out of open data [40]. The recognition that open data acts as a catalyst, enabler, and foundation for entrepreneurial activity has been recognized both in policy circles [41] and in academic scholarship [22, 39]. At the same time, the real economic value contribution from data comes from the business insights that are extracted from it [42]. Creative business insight falls to entrepreneurs who, in their quest to carve out niche markets and diversify revenue streams, find themselves best placed to bring together information and knowledge from different sources [40].

Promoting Innovation: Open data has been seen to have a significant positive impact on the creation of economic value by supporting increased rates of innovation [43, 44]. The use of open data creates multiple business opportunities for companies: both established firms [44], as well as SMEs and entrepreneurs [39], can use open data in combination with new technologies and existing proprietary information to develop new products and services, improve existing offerings, and drive forms of innovative entrepreneurship that are at once commercially profitable and beneficial for the public good [22]. The full business potential of open data cannot, however, be harnessed if governments withhold access to particular datasets or privilege the release of certain types over others [45].

Impact on Economic Efficiency and Resource Allocation: The provision of open data by government and private organisations creates economic efficiency gains by enabling the better, more responsible allocation of existing resources and by minimizing resource wastage [45, 46]. The provision of open data cuts data processing and transaction costs, facilitates strategic connections between economic entities, and

empowers key actors [47]. Ready access to data can also reduce asymmetries of information [40].

Open Data Business Models: The widespread availability of data, and especially the increasing prevalence of open government data, has resulted, not only in the establishment of new companies, but in the emergence and adoption of new data-centric business models to drive product development, service provision and expansion into new markets [48–50]. These so-called 'Open Data Business Models' or ODBMs [49] place open data and associated tools at the heart of strategic decision-making and the development of goods and services, and are essential for firms wanting to leverage the economic potential of open data.

4.3 Open Data for Promoting Good Governance

Impact on Public Accountability: Open data can be a powerful enabler for public accountability, whereby the publishing of government data online in machine-readable and easily interpretable formats for consumption by citizens and firms effectively holds up government actions to general scrutiny [51, 52]. Similarly, the shift towards the provision of big open data by government has made the imperative for public accountability even greater [53]. Open government portals need, therefore, be designed to fully support ordinary citizens engaged in public accountability efforts [54].

Impact on Transparency: The opening of data has profound implications for notions of public transparency [55, 56]. Public transparency is one of the three principles around which the open government paradigm has been constructed, and a major objective of releasing open data for governments has become its promotion [53]. However, there is inconclusive evidence that the conscious opening of publicly held data will automatically lead to increased transparency [4]. Instead, progression from an open government to a transparent one is not considered altogether straightforward [56], as it is argued that the release of open data in machine-readable formats is seldom accompanied by the mechanisms necessary to facilitate its better understanding [55].

Creating Public Sector Efficiency Gains: Open data has the potential to modernize and streamline public sector operations; especially when used in combination with technology, and released in line with appropriate regulatory frameworks [57]. In fact, the improvement of operational efficiency is often listed as a compelling reason for governments to open their data to the public [24, 58]. Further, the realization of public sector efficiency gains through the release of open data is sometimes considered a precondition for good governance [59].

Open Data and Civil Society Organisations: The release of open data has been seen to have a positive impact on the day-to-day operations of civil society organisations [40], and has the potential to transform the role that these actors play in the governance process [60]. Detailed information on the location of resources, and on the quality of existing systems, can both facilitate the resource mobilization activities performed by this sector [40] and support dialogue or collective deliberation with public officials [61]. Equipped with this knowledge, civil society organisations are also well placed to

create awareness about open data and to identify those datasets which possess intrinsic social value [60]. However, it may be argued that while open datasets support civil society bodies, the interests of marginalised groups represented by these organisations may not be immediately served by the provision of machine-readable data online [61].

5 Discussion: Insights and Implications for Policymaking

In this section, we aim to propose some practical recommendations for open government data implementation. In general, open data is seen to play a significant role in the generation of social capital and creation of value within the economic context. Open data can also contribute to the advancement of good governance. These aspects imply that widespread access to data offers significant opportunities for policymakers to develop better-informed public policy, and can transform governance processes by encouraging participatory and collaborative decision-making. The maximization of the benefits accruing from open data in the move towards open governance and the provision of inputs for evidence-based policy making will be discussed below.

5.1 Exploring the Implications of Open Data-Driven Transformation

Digital transformation is changing the process of policymaking and altering governance models in a disruptive way. Developments in open data, data processing, data mining and visualizations combined with social media, participatory tools and civic engagement are responsible for the changes in the policy-making field [1]. Our research has highlighted six pre-requisites for open data-driven transformation, mapped in Fig. 1:

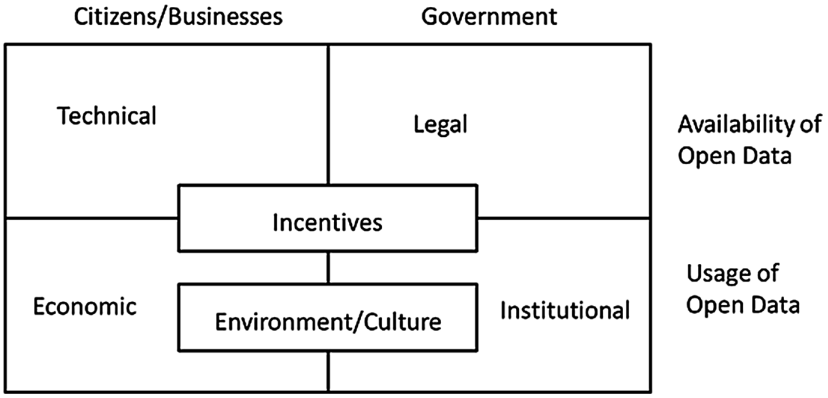

Fig. 1. Mapping open data-driven transformation

1. *Technical*: Facilitating user-friendly access to high quality machine-readable data.
2. *Legal*: Creation of an appropriate legal and regulatory framework for the release of open data.

3. *Environmental*: Fostering of a permissive culture that supports government-citizen collaboration, co-creation and, eventually, open innovation.
4. *Institutional*: Streamlining of government organisations to better engage with open data and convert basic information into public goods.
5. *Economic*: Adoption of data-centric business models for strategic decision making and product/service development.
6. *Financial*: Incentivisation of data-driven actions leading to innovation.

In order to maximize the benefits of open data in the move towards open governance, data needs to be open, the re-use of data should be facilitated to achieve social and economic value, interdisciplinary collaboration should be legitimized, as well as participation and engagement practices in decision making encouraged.

5.2 Towards Open Governance

Open governance is a concept that has emerged from the understanding that information belongs to the public, and, alongside with electronic democracy practices, can transform electronic government to electronic governance [62]. Open government is highly related to the collaborative governance concept, since open data increases the possibilities for knowledge development, decision making and interdisciplinary collaboration [63]. Citizens' willingness to engage is related to both their perception of government openness to integrate public opinion to formulate decisions and the amount of open public sector information to which they have access [13]. When analysing collaborative governance, [64] identified that information sharing and cooperation are the main elements framing the use of ICT to enable collaborative governance along with participation and engagement practices in decision making. In this sense, collaboration in government has as its main goal the enabling of stakeholders to participate in decision making processes that are efficient and effective [15].

6 Conclusion

The objective of this study was to explore the impact that the access to, and free usage of, stored (mainly public sector) data has on society, the economy, and on good governance. In order to achieve this aim, our research made use of a systematic review of current scholarship, and developed a framework that sets out the implications of this new paradigm for policymakers within modern-day governments. This study identified 12 key domains within which the use of open data had the potential to transform and generate social, political, and economic value. These domains were then used as the basis for the development of the foundations for policy recommendations in the field of open data; most particularly to highlight the contribution made by open data to transformations in government decision-making and policymaking.

The results of the literature review highlighted several key insights, pre-conditions for effective open data usage. The first of these is that open data has become a critical input for evidence-based policy-making, which can be achieved through new tools and technologies such as big data analytics. However, data needs to be not just available but

accessible in formats that re-usable to generate social, political and economic value. Third, the opening up of data increases its re-use as it exposes information to a wider audience. However, data in circulation needs to be of a certain high quality, and needs to be accompanied by the right tools and technology for it to reach those who can actually make use of it. Fourth, there is a need to invest in technology and skills that will create value out of open data. Fifth, open data may be free but data analytics still requires an investment. Money, time and effort need to be spent to maintain and sustain open data portals. And finally, there is requirement to open data across the board whilst setting out strategic incentives to channel value generation to where it is needed.

In conclusion, it is vital that not only is data available and easily accessible to the public, but also that the appropriate tools and regulatory frameworks are put in place to ensure that the use and re-use of open data is meaningful.

References

1. Janssen, M., Wimmer, M.A.: Introduction to policy-making in the digital age. In: Janssen, M., Wimmer, M.A., Deljoo, A. (eds.) Policy Practice and Digital Science. PAIT, vol. 10, pp. 1–14. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-12784-2_1
2. Hossain, M.A., Dwivedi, Y.K., Rana, N.P.: State-of-the-art in open data research: insights from existing literature and a research agenda. J. Organ. Comput. Electron. Commer. **26**(1–2), 14–40 (2016)
3. Kučera, J., Chlapek, D., Nečaský, M.: Open government data catalogs: current approaches and quality perspective. In: Ko, A., Leitner, C., Leitold, H., Prosser, A. (eds.) Technology-Enabled Innovation for Democracy, Government and Governance - International Conference on Electronic Government and the Information Systems Perspective, pp. 152–166. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40160-2_13
4. Janssen, M., Matheus, R., Longo, J., Weerakkody, V.: Transparency-by-design as a foundation for open government. Transform. Gov. People Process Policy **11**(1), 2–8 (2017)
5. Veljković, N., Bogdanović-Dinić, S., Stoimenov, L.: Benchmarking open government: an open data perspective. Gov. Inf. Q. **31**(2), 278–290 (2014)
6. Bauer, F., Kaltenböck, M.: Linked Open Data: The Essentials. Edition mono/monochrom, Vienna (2011)
7. Ruijer, E., Grimmelikhuijsen, S., Meijer, A.: Open data for democracy: developing a theoretical framework for open data use. Gov. Inf. Q. **34**(1), 45–52 (2017)
8. Safarov, I., Meijer, A., Grimmelikhuijsen, S.: Utilization of open government data: a systematic literature review of types, conditions, effects and users. Inf. Polity **22**(1), 1–24 (2017)
9. OECD: Emerging issues: the internet of things. In: OECD Digital Economy Outlook 2015. OECD Publishing, Paris (2015)
10. Zuiderwijk, A., Janssen, M., Davis, C.: Innovation with open data: essential elements of open data ecosystems. Inf. Polity **19**(1/2), 17–33 (2014)
11. Danneels, L., Viaene, S., Van den Bergh, J.: Open data platforms: discussing alternative knowledge epistemologies. Gov. Inf. Q. **34**(3), 365–378 (2017)
12. Gonzalez-Zapata, F., Heeks, R.: The multiple meanings of open government data: understanding different stakeholders and their perspectives. Gov. Inf. Q. **32**(4), 441–452 (2015)

13. Mellouli, S., Luna-Reyes, L.F., Zhang, J.: Smart government, citizen participation and open data. Inf. Polity **19**(1, 2), 1–4 (2014)
14. Viale Pereira, G., Cunha, M.A., Lampoltshammer, T.J., Parycek, P., Testa, M.G.: Increasing collaboration and participation in smart city governance: a cross-case analysis of smart city initiatives. Inf. Technol. Dev. **23**(2), 1–28 (2017)
15. Attard, J., Orlandi, F., Scerri, S., Auer, S.: A systematic review of open government data initiatives. Gov. Inf. Q. **32**(4), 399–418 (2015)
16. Kitchenham, B.: Procedures for performing systematic reviews. Joint Technical Report – Technical Report TR/SE-0401& NICTA Technical Report 0400011T.1, Keele University, June 2004 (2004)
17. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **26**(2), XIII–XXIII (2002)
18. OECD: Rebooting public service delivery: how can open government data help to drive innovation? An OECD comparative study (2017). http://www.oecd.org/gov/Rebooting-Public-Service-Delivery-How-can-Open-Government-Data-help-to-drive-Innovation.pdf. Accessed 15 Sept 2017
19. Attard, J., Orlandi, F., Auer, S.: Data driven governments: creating value through open government data. In: Hameurlain, A., Kung, J., Wagner, R., Anjomshoaa, A., Hung, P.C.K., Kalisch, D. (eds.) Transactions on Large-Scale Data-and Knowledge-Centered Systems XXVII - Special Issue on Big Data for Complex Urban Systems, pp. 84–110. Springer, Berlin Heidelberg (2016). https://doi.org/10.1007/978-3-662-53416-8_6
20. Daglio, M., Gerson, D., Kitchen, H.: Building organisational capacity for public sector innovation. In: Background Paper Prepared for the OECD Conference "Innovating the Public Sector: From Ideas to Impact", 12–13 November 2014, Paris (2015)
21. Linders, D., Wilson, S.C., Bertot, J.C.: Open government as a vehicle for government transformation. In: Weerakoddy, V., Reddick, C.G. (eds.) Public Sector Transformation through E-Government: Experiences from Europe and North America, pp. 9–23. Routledge Press, New York (2012)
22. Jetzek, T., Avital, M., Bjorn-Andersen, N.: Data-driven innovation through open government data. J. Theor. Appl. Electron. Commer. Res. **9**(2), 100–120 (2014)
23. Chan, C.M.L: From open data to open innovation strategies: creating e-services using open government data. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), Proceedings, pp. 1890–1899. IEEE Press (2013)
24. Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., Dwivedi, Y.K.: Open data and its usability: an empirical view from the citizen's perspective. Inf. Syst. Front. **19**(2), 285–300 (2017)
25. Ruijer, E., Grimmelikhuijsen, S., Hogan, M., Enzerink, S., Ojo, A., Meijer, A.: Connecting societal issues, users and data. Scenario-based design of open data platforms. Gov. Inf. Q. **34** (3), 470–480 (2017)
26. Craveiro, G.S., Machado, J.A.S., Machado, J.S.: The use of open government data to citizen empowerment. In: ICEGOV 2015–2016 Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance, pp. 398–399. ACM Press, New York (2016)
27. Geiger, C.P., Von Lucke, J.: Open government and (linked) (open) (government) (data). JeDEM-eJ. eDemocr. Open Gov. **4**(2), 265–278 (2012)
28. Maier-Rabler, U., Huber, S.: "Open": the changing relation between citizens, public administration, and political authority. JeDEM-eJ. eDemocr. Open Gov. **3**(2), 182–191 (2011)

29. Johansson, D., Lassinantti, J., Wiberg, M.: Mobile e-services and open data in e-government processes: transforming citizen involvement. In: Proceedings of the 17th International Conference on Information Integration and Web-Based Applications & Services (iiWAS 2015). ACM, New York (2015). Article No. 8
30. Zuiderwijk, A., Janssen, M.: Open data policies, their implementation and impact: a framework for comparison. Gov. Inf. Q. **31**(1), 17–29 (2014)
31. Conradie, P., Choenni, S.: On the barriers for local government releasing open data. Gov. Inf. Q. **31**(Supplement 1), S10–S17 (2014)
32. Kassen, M.: A promising phenomenon of open data: a case study of the Chicago open data project. Gov. Inf. Q. **30**(4), 508–513 (2013)
33. Hardy, K., Maurushat, A.: Opening up government data for big data analysis and public benefit. Comput. Law Secur. Rev. **33**(1), 30–37 (2017)
34. van Zoonen, L.: Privacy concerns in smart cities. Gov. Inf. Q. **33**(3), 472–480 (2016)
35. Scassa, T.: Privacy and open government. Future Internet **6**(2), 397–413 (2014)
36. Floridi, L.: Open data, data protection, and group privacy. Philos. Technol. **27**(1), 1–3 (2014)
37. Meijer, R., Choenni, S., Alibaks, R.S., Conradie, P.: Bridging the contradictions of open data. In: Castelnovo, W., Ferrari, E. (eds.) Proceedings of the 13th European Conference on E-government, pp. 329–336. Academic Conferences and Publishing International Limited, Reading (2013)
38. Walker, J., Simperl, E.: Analytical Report No. 10: Open Data and Entrepreneurship. Report for the European Data Portal. https://www.europeandataportal.eu/sites/default/files/analytical_report_10_open_data_and_entrepreneurship.pdf. Accessed 07 Mar 2018
39. Lakomaa, E., Kallberg, J.: Open data as a foundation for innovation: the enabling effect of free public sector information for entrepreneurs. IEEE Access **2013**, 558–563 (2013)
40. Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., Doshi, E.A.: Open data: unlocking innovation and performance with liquid information. Report by the McKinsey Global Institute, The McKinsey Centre for Government, and the McKinsey Business Technology Office, October 2013 (2013)
41. Meyer, E.: The impact of open data. The White House – Blog, 08 April 2014 (2014). https://obamawhitehouse.archives.gov/blog/2014/04/08/impact-open-data. Accessed 20 Sept 2017
42. Kamal, I.: Metrics are easy: insight is hard. Harvard Business Review, 24 September 2012 (2012). https://hbr.org/2012/09/metrics-are-easy-insights-are-hard. Accessed 20 Sept 2017
43. Susha, I., Grönlund, A., Janssen, M.: Driving factors of service innovation using open government data: an exploratory study of entrepreneurs in two countries. Inf. Polity **20**(1), 19–34 (2015)
44. Zimmermann, H.-D., Pucihar, A.: Open innovation, open data and new business models, 1 September 2015 (2015). https://ssrn.com/abstract=2660692. Accessed 21 Sept 2017
45. Carrara, W., Chan, W.S., Fischer, S., van Steenbergen, E.: Creating Value through Open Data. Study prepared as part of the European Data Portal (2015)
46. European Data Portal: Benefits of open data (2017). https://www.europeandataportal.eu/en/using-data/benefits-of-open-data. Accessed 21 Sept 2017
47. Berends, J., Carrara, W., Radu, C.: Analytical Report No. 9: The Economic Benefits of Open Data. Report for the European Data Portal. https://www.europeandataportal.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf. Accessed 07 Mar 2018
48. Koski, H.: The impact of open data: a preliminary study. Ministry of Finance Publications 15b/2015, March 2015. Government of Finland (2015)
49. Zeleti, F.A., Ojo, A., Curry, E.: Exploring the economic value of open government data. Gov. Inf. Q. **33**(3), 535–551 (2016)
50. Bonina, C.M.: New business models and the value of open data: definitions, challenges and opportunities. In: New Economic Models in the Digital Economy, London, UK (2013)

51. Misuraca, G., Viscusi, G.: Is open data enough? E-governance challenges for open government. Int. J. Electron. Gov. Res. **10**(1), 18–34 (2014)
52. Guéret, C., de Boer, V., Schlobach, S.: Let's "downscale" linked data. IEEE Internet Comput. **18**(2), 70–73 (2014)
53. Sivarajah, U., et al.: The role of e-participation and open data in evidence-based policy decision making in local government. J. Organ. Comput. Electron. Commer. **26**(1–2), 64–79 (2016)
54. Lourenço, R.P.: An analysis of open government portals: a perspective of transparency for accountability. Gov. Inf. Q. **32**(3), 323–332 (2015)
55. Temiz, S., Brown, T.: Open data project for e-government: case study of Stockholm open data project. Int. J. Electron. Gov. **9**(1–2), 55–84 (2017)
56. Kornberger, M., Meyer, R.E., Brandtner, C., Höllerer, M.A.: When bureaucracy meets the crowd: studying "open government" in the Vienna city administration. Organ. Stud. **38**(2), 179–200 (2017)
57. Donker, F.W., Van Loenen, B., Bregt, A.K.: Open data and beyond. ISPRS Int. J. Geo-Inf. **5** (4), 48 (2016)
58. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. **29**(4), 258–268 (2012)
59. Graves, A., Hendler, J.: A study on the use of visualizations for open government data. Inf. Polity **19**(1–2), 73–91 (2014)
60. National Democratic Institute (NDI): Open data is only the beginning: civil society organisations as public data analysts (2013) https://www.ndi.org/sites/default/files/2013GhanaStanfordTransparencyPaper.pdf. Accessed 20 Sept 2017
61. Meng, A.: Investigating the roots of open data's social impact. JeDEM-eJ. eDemocr. Open Gov. **6**(1), 1–13 (2014)
62. Klaus, L.C.O.: Transforming armed forces through military transparency: open government challenges in a world of secrecy. Transform. Gov. People Process Policy **10**(1), 99–119 (2016)
63. Kamateri, E., et al.: A comparative analysis of tools and technologies for policy making. In: Janssen, M., Wimmer, Maria A., Deljoo, A. (eds.) Policy Practice and Digital Science. PAIT, vol. 10, pp. 125–156. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-12784-2_7
64. Open Knowledge Definition (2017). http://opendefinition.org/okd/. Accessed 20 Sept 2017

Towards Open Data Quality Improvements Based on Root Cause Analysis of Quality Issues

Csaba Csáki(✉)

Corvinus University of Budapest, Budapest, Hungary
Csaki.Csaba@uni-corvinus.hu

Abstract. Commercial reuse of open government data in value added services has gained a lot of interest both as practice and as a research topic over the last few years. However, utilizing open data without proper understanding of potential quality issues carries the risk of undermining the value of the service that relies on public sector information. Instead of establishing a data quality assessment framework this research considers a review of typical open data quality issues and intends to connect them to the causes leading to these various data problems. Open data specific problems are concluded from a case study and then theoretical and empirical arguments are used to connect them to root causes emerging from the peculiarities of the public sector data management process. This way both practitioners could be more conscious about appropriate cleansing methods and participants shaping the data management process could aim at eliminating root causes of data quality issues.

Keywords: Open data · Open data reuse · Data quality framework
Open data quality · Public procurement · Data cleansing · Root cause analysis

1 Introduction

While the idea of open government data (OGD) has been around for some time, every few years there is a rejuvenated scientific interest in the topic. The motivation had been changing from promoting accountability and transparency to supporting e-Government to the push for open government. The latest trend is based on economic interest, namely the idea of innovative, commercial reuse of public sector information (PSI) [14]. The changing focus came with changing research goals and shifts in research objectives.

One of the main conditions of successful reuse is quality data [6]. It is no surprise, therefore, that over the years a lot of scientific and on-the-field efforts have been reported to address the issue. One well-quoted are deals with the question of provenance [21] with the leading idea covering the maturity level of linked open data published [2]. However, the shortcoming of the five star model is that it focuses almost exclusively on the area of linked open data, thus only addresses the related subset of quality features (such as traceability, linking standards, and machine readability). There are ample results that offer frameworks dealing with a wider range of quality dimensions – both in general [17, 32] or focusing on OGD in particular [10] –, but these approaches mostly cater for the assessment of datasets [1, 30, 35]. While the ability to

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 208–220, 2018.
https://doi.org/10.1007/978-3-319-98690-6_18

judge the quality of a given data set is useful, typically there are no guidelines how to address these issues – neither how to eliminate them on the producer side, nor what to do about them before (secondary) utilization. While considering the outcome could improve certain aspects of structuring and storing data, existing frameworks are unable to consider the bigger picture of policies and regulations or the organizational context [36]. So, although there are ample proposals about organizational preparation or guidelines considering technology, even these best practices do not seem to be enough regarding quality of the content [37]. The assumption here is that guidelines related to various aspects of quality are no help mainly because they do not focus (enough) on the root causes of ODG quality issues. Thus while there is considerable research efforts addressing certain specific areas related to producing good quality open data, relatively less attention is paid to why data quality issues are still present despite the best efforts and available experience.

The research reported here was focusing mainly on the latter problem and its objective was to understand the root causes of quality issues affecting the publication of PSI intended for reuse. This paper is organized as follows: the next section reviews frameworks of open data quality and considers characteristics of the public sector as context. The methodology section raises research questions and presents the methodology followed. Then the attention is turned to a review of public sector data quality issues through a case study. This is followed by a discussion of potential root causes of issues identified along with their generalization. The paper closes with conclusions and the usual recommendations of future research ideas.

2 Open Government Data and Related Quality Considerations

2.1 The Push for Open Government Data

The term open data as a popular concept was first used in 2006 when the Open Knowledge Foundation (OKF) has announced its Open Knowledge Definition [5, 18]. This was a general call for opening up scientific and other data for further use or reuse. There are other forces behind publishing public sector data, however, some of which are rooted in early civil society movements, albeit the dawn of the Internet and related technologies have also played a major role in an increased interest (by providing new possibilities). Meeting integrity and accountability goals in democratic societies is anchored by transparency that in turn assumes a requisite level of openness whereby non-government actors (the public) have mechanisms to know what governmental actors are doing [4]. Thus, data about governmental behavior may be used to hold actors of the public sphere to account [20]. Over time the goal of promoting accountability and transparency has been overtaken by the motivation to use technology in support of e-Government initiatives [12], and later by the push for open government [34] also fueled mostly by advances in technology. In the former technology was an enabler of open data, while in the context of open government open data became the main piece of the puzzle as an enabler of advanced participative democracy [28]. However, the latest push for publishing data generated or controlled by public

actors considers commercial utilization of such data. Indeed, reusing public sector information in innovative value added services is turning into a serious market [19].

Since more and more data and information generated in various public policy domains are being captured, digitized, and stored, it would be difficult (at least in democratic regimes) to argue for completely shielding such digital records from public scrutiny on the one hand or potential utilization on the other. While users of public or open government data are assumed to be able to utilize them with ease, the fact is that open data is not free of quality issues [3]. The higher demand for more open data does not necessarily come with increased quality. Assessing the quality of data in general or open data in particular is not a straightforward exercise.

2.2 Information Quality Dimensions

There are different frameworks allowing for discussion over data or information quality (DQ/IQ) and within which DQ/IQ may be assessed (for simplification we do not enter into a theoretical discussion about the difference of data and information, we simply regard information as data in use, as that fits the problem at hand). The technical view associates quality with the accuracy of the data in products such as databases looking at timeliness of update, system reliability, system accessibility, system usability and system security [16]. Another, the machine readability approach [35] is concerned with linking, finding, relating and reading data typically using automated processes, and characteristics usually considered include number of formats, traceability, automated tracking, use of standards, or provenance. Perhaps the most commonly used simple definition of user side information quality interprets the term as "fit-for-use" [31]. However, IQ defined this way remains a relative construct whereby data considered appropriate for a given use may not display acceptable attributes in another setting [25]. Furthermore, fit-for-use does not immediately allow for ready measurability and it requires additional detail in order to be operationalized [11].

When it comes to actually assessing DQ/IQ, it is typically related to "a set of 'dimensions' that are usually defined as quality properties or characteristics" ([22] p. 2). However, this approach has led to a proliferation of features and dimensions – as various models proposed distinguished sixteen [15], twenty-eight [7], or even thirty-two [29] different dimensions – although the most important dimensions appear to overlap. It is possible to organize these features along a natural timeline of the steps normally taken when exploring new data: Awareness and availability, Accessibility, Readability, Technical qualities of the data, Content and structure, Traceability, Usability, Fit-for-purpose – each in turn including several sub-features (depending on the framework). However, the quality of the data as stored, accessed and manipulated can substantially differ from the quality of the information that the data contains or that the data can offer in terms of information gleaned from it.

2.3 Some Special Characteristics of the Public Sector

As an important step towards understanding why quality issues may happen in public context, it is worthwhile to look at a few fundamental characteristics of the public sphere. One essential difference between the private and the public sector is that public

policies are created based on public interest while private corporate goals serve private interest. Public value cannot be defined by commercial categories only, and governments thus have responsibilities related to fulfilling non-commercial goals which in turn increases costs [9]. In addition, governmental choices have lasting, long-reaching effects. Indeed, the policy making function creates the – formal and legal – environment within which society and economy operates. [23] points out that the notion of public interest does require some form of sympathy with the needs of others which, therefore, may not be reconciled with the market notion of maximizing economic opportunities and personal wealth.

Regarding data management, governmental functions are guided by laws and regulations influencing related processes, tasks, roles, and responsibilities. Furthermore, there are specific regulations controlling the release of data (typically in the context of the so called 'right to information' law). Thus, utilizing data from public websites presumes some level of legitimacy on the part of the immediate publisher. This setting has an impact on the way of producing and collecting data as well as the way open data is generated from the data stored. Information collection is usually done through forms defined by the corresponding law such as a relevant act. All in all, these result in a data lifecycle that is different from its counterpart in the private sector: the data does not connect, rather open data as published sharply separates the supplier from the consumer. In addition, changing any part of the data producing side would require changes in the corresponding regulatory component - which might take time (due to the formal processes involved).

2.4 Open Government Data Quality Frameworks

Despite the difference in context, open data specific 'quality frameworks' often offer similar categorizations. One ODQ stream considers technical standards and abilities as well as processes and outcomes of producing and managing datasets, but also considers the timeliness of data (i.e. whether it is out-of-date). Another stream is centered on the availability and accessibility of various types of data or data in certain categories, while also measures whether intended audiences are aware of the availability of relevant datasets and if data is easy to find [5]. Yet another set of frameworks is concerned about specific sectors and take into account the content of the datasets. Finally, it is customary to ask about the value of open data, which, in general terms considers the needs of end users [11]. However, irrespective of the approach, there is a tendency here too to favor characteristics that are measurable. The current disposition of open data quality characteristics is aptly demonstrated by [35] which, in pursuit of the measurability of ODQ, define and operationalize 68 metrics along 6 dimensions. Most of these characteristics do not differ from 'regular' data quality dimensions, although there are some additional concerns related to access to and freshness of data – as well as to potential fees charged. However, irrespective of the assessment methods, there are still typical problems with public data made available. Therefore, the objective of this research was to understand and map the root causes why (and where) certain types of public open data quality issues and errors happen – with the ultimate intent to make recommendations what to do about them.

3 Methodology: Theoretical Arguments with a Case Study

Under the above objective the following research questions are proposed: (1) What is the status of the quality of Open Government Data from the point of view of 'content'? (2) What are the main reasons (root causes) behind OD content issues in public context? (3) What to do about improving content quality of OGD? This paper focuses on the second question, using the first as support – but would not have the room to address the last question here.

To demonstrate typical content errors in public sector data sets along with their causes a case study methodology was designed [33]. To explore the reasons behind the quality issues of open government data the research plan contained the following steps: (1) review typical data quality issues (that impact reuse); (2) illustrate them through examples from the case; (3) use the case study to identify potential causes of errors; (4) propose a generalization of those root causes.

This was an intrinsic [24], single case [8] research study, where the exploration of the case (i.e. data collection) involved (a) investigating a complex open data set; (b) reading documents describing the data set (including its structure and known issues); and (c) email communication with a representative of the issuer of the data for further clarification. To establish root causes of issues presented in the case theoretical arguments from relevant literature were applied. Finally, in the last step the generalization was based on the understanding of the immediate context and process of producing open government data (concluded from the literature review).

The data set used (as the case) was the public procurement (PP) open data of the European Union (EU), selected based on its special characteristics regarding size, complexity, regulatory context and multiple stakeholders. Counterarguments may be that the EU PP legal context is complex and further burdened by a multinational setting. However, while understanding the depth of the case might be a challenge for some readers outside PP, the descriptive power of the case well offsets the efforts required. As part of its broader e-Government initiative the European Commission (EC) has been an advocate of the open data movement for some time. "The European Data Portal" (https://www.europeandataportal.eu/) offers public sector information originated in the member states and portal data may also be repurposed. Through the Directive 2003/98/EC the EC has set up the legal framework to allow the reuse of public sector information. One key component of the EU Open Data initiative (https://ec.europa.eu/digital-single-market/en/open-dat) is the Tenders Electronic Daily dataset comprised of public procurement data of the twenty-eight member states. While the data is accessible as part of a daily journal (the online version of "*Supplement 32 to the Official Journal of the European Union*"), there is an annual release of summarized historical data in CSV format (dating back to 2006 at https://data.europa.eu/euodp/en/data/dataset/ted-csv). Data covers purchases of public procurement that fall above given threshold amounts stipulated in EU regulations for procurement. Other than EU members, affiliated countries also publish tender and award notices in the TED Journal to gain access to the EU market. Data in the Journal are collected from standardized public procurement forms as required by the corresponding EU Directives (2014/17 and/18) and their Annexes. The data originally recorded store information captured

from the contract notices reported in standard forms #2, #4, #5, or #17. These forms announce information concerning a future purchase (i.e. call for tender). Another set of data covers contract award notice information on the outcomes of the procurement obtained from standard forms #3, #6 or #18. Data is entered through online versions of these forms, one notice at a time. The open datasets published annually come with a codebook [26] describing the fields in the files made available. In addition, for advanced users of the CSV datasets a user guide is available [27] providing information about known issues and difficulties.

The TED open data is very complex because the CSV data files have three levels of procurement information embedded: (a) contract notices (CN); (b) contract award notices (CAN); and (c) contract awards (CA) (the last two published in one file, of course). While the process of public procurement is inherently complicated, for now it should suffice to state that one and occasionally two CNs lead to one CAN (this is because a CN may have a preliminary notice with a separate CN ID), but one CAN may lead to one or more CAs associated with it (as a single call may have several parts or lots with each leading to a separate contract being awarded under the same CAN ID but individual award CA IDs). Issuers of notices are called "contracting authority". Each annual dataset is published in CSV format using UTF-8 coding. All data files were (first) downloaded January 17, 2017. There were two types of data – notices and awards – from 2009 to 2015 (the first three years had to be omitted), and the size of the sixteen files was over 2 GB (each ranging from 130 to 280 MB). MS Excel and MS Access (both from Office 2010 on Win7 OS) have been utilized to open and investigate the structure and content of the files. In addition, SPSS (v22.2) and Oracle Database (11g r. 11.2.0.4) were also used to investigate data quality.

4 Typical Quality Issues in the Case – and Their Root Causes

While access related features (Availability, Accessibility and Readability as well as Traceability) are important as a starting point for OD utilization, they are less relevant in our context of actual reuse. On the other hand, content related characteristics (such as Technical qualities, Content and structure, as well as Fit-for-purpose) are main concerns that immediately influence usability. According to literature, data content errors are typically organized into four categories (based on [25]): Missing data (missing field or missing value); Duplication (physical duplication or logical duplication); Error with meaning or interpretation (syntactical error, out of bounds, format error, data does not make sense in context); and Inconsistency (inconsistency between data fields, data tables, databases or outside sources).

Case Problem: Successful opening of the file(s) is followed by the investigation of the Technical qualities of the data. Due to the nature of CSV, the original dataset as published does not carry data types. The typical result is formatting errors. Even in Excel – the tool users would use to open CSV –, fields containing data that look like (calendar) date would indeed be interpreted as calendar date, resulting in automatic corrections, which are often faulty (e.g. 2004 may become 2004 January 1).

Root Cause Analysis: These issues are related to the process of generating open data and the publishing format being used. For example, a formatting error may be the result of inadequate consideration and lack of flexibility in data formats (especially in international, multi-language context). Also, lack of data type information in simple standard formats may lead to misinterpretation by more sophisticated tools.

Case Problem: Since EU members may use any of the official languages for their PP announcements, basic UTF-8 reading with a default language (such as English) resulted in scrambled characters for languages like Greek, Hungarian, Swedish, etc. Interestingly, each tool used had its own way of dealing with this problem: SPSS, and Oracle could only read setting of one language or another, MS Access had an "all" setting for UTF-8 font mixing, allowing for text from every EU language to be displayed properly, while MS Excel required the "import" function for proper UTF-8.

Root Cause Analysis: It appears that CSV does not carry language information, but UTF-8 requires a so called BOM character for font mixing.

Case Problem: The case data files had a lot of text fields, some of them are quite long – and for most tools the length of textual data is an issue: some truncate lines while others simply drop whole records with fields of inappropriate sizes. So, as the result of the above, the actual data as opened may have missing fields, missing content, or inappropriate content or even inconsistencies within the dataset.

Root Cause Analysis: The loss of information is due to technical issues such as the lack of data type information or the use of long text in one field.

Case Problem: The most important step (before any use), is the checking of content and its validity. The outcome of a procedure (CN) may be a successful award (CAN with one or more CA), modification, cancellation, or cancellation with a new call. Unfortunately, cancellation and modification information are not always recorded properly (cancellation or modification flag is missing from the form) leading to either missing information or duplicate records. As a result, the CSV output generated had missing flags and duplicate CN IDs.

Root Cause Analysis: The cause of such problems is rooted in the mode of entering data into the forms, especially online: (a) the forms themselves could be faulty (such as having missing fields); (b) there could be human error (using the wrong form or lack of knowledge about how to fill out the form) on part of the contracting authority personnel; and (c) these may be combined with inappropriate sanity check or lack thereof. In addition, (d) the algorithm generating the OD output file may be misled by the inappropriate information.

Case Problem: Another form of information loss happened when there were multiple values in one field and most tools could not separate them. This happened in two ways: (a) when there were two winners to be announced, instead of two separate award (CA) IDs the name of both winners were entered into the corresponding field; (b) categorization of the product to be purchased is based on so called CPV codes, but complex purchases may require one main and several secondary CPVs.

Root Cause Analysis: Situation "(a)" is clearly a human error; while "(b)" relates to the way forms are defined (instead of allowing for recursion, repeated values are entered into the same field using some separators). The latter issue causes a problem either way when output records are generated.

Case Problem: There were duplicate lines where certain CN IDs were erroneously coupled with CAN IDs from other calls. While CANs may appear several times in case of multiple awards for one call, calls (a given CN ID) should not be repeated.

Root Cause Analysis: This appears to be a CSV generation issue, as checking such duplicates on the TED search page returns only one item for a given CN ID. Furthermore, in 2014 there was a change in forms – and generating the CSV data from data captured using the old forms were executed according to new forms leading to irregular duplications (which could have easily been filtered out).

Case Problem: There were inconsistent values: each type of call should use the corresponding form, however, there are a reasonable number of records where the form number in the record does not match the type of call.

Root Cause Analysis: This is a data validation issue during the submission of the form (likely coupled with human error).

5 Generalization of the Causes Behind OGD Quality Issues

It follows from the nature of public sector activities, namely that they are governed by policy (with underlying strategies) and corresponding laws and regulations, that legal foundation for publishing OGD could already have an influence on the data that may, must, or should not be released and how they were supposed to be published [13]. The legal frame controls what may be published, in what format or by whom. This carries a certain risk of errors when it comes to content and format of data being made openly accessible.

Implementation of the regulations poses a challenge as well – organizationally, process-wise and regarding technological support. In the case presented, the EC directives stipulate that the collecting and entering of data is organized around filling out specific forms. These are not always on-line, thus entering data online often means copying from hand-filled forms. This is a major source of typos and errors. Even with online forms there is a possibility of inappropriate completion of data fields – some of which are deliberate [27]. Individual behavior and lack of control mechanisms built-in when uploading data using the forms will eventually lead to error in generating the output format from data stored. Fields may be missing from the form, data is not even entered into the form (field left empty). Even if data was entered, often the data is a dummy value just to fill in the field (to avoid being caught by validation if the field is empty). Allowing multiple values in the same field is a serious form issue, resulting in serious challenges during statistical analysis.

Understanding the meaning of various fields requires in depth knowledge not only of public procurement in general but specific details of EU procedures, including the intent and use of various forms. For example, fields in the csv files did not fully reflect

either the fields in the TED DB (presented as documents through an online interface) or the original forms contracting authorities required to use when submitting data related to calls and results. This is not unusual for public sector data collection typically based on forms. Based on the analysis in the previous section and on the understanding of the role of forms in the public data management process, Table 1 summarizes the generalization of root causes identified.

Table 1. Overview of issues and causes – through examples

Type	Case examples	Root cause	Reasons generalized
Format issues	Date is not interpreted properly	Representation and data type issue	Either in the DB or during generation of the open version inconsistent formatting is used – and most often data type information is lost
	Scrambled characters appeared for certain countries	UTF-8 does require BOM for font mixing	Machine readability of even standard forms have language dependencies
Missing data	No indication of cancellation	Form error	Public sphere data collection forms are part of the regulation but often are out of sync with the process
		Data entering error	Due to the complexity of legally controlled processes, mistakes are easy to make
		Error in checking the validity of filling out the form	Checking relationships between data being entered and data in the DB is not straightforward in this context
Logical duplication	CNs are mixed with CAN belonging to a different procedure	Output generation error: During the generation of the open version (CSV), records were connected inappropriately	Data recorded in form (using online of pdf) are then stored in various databases and the open version is generated using a dedicated process (and algorithm) which may introduce errors

(*continued*)

Table 1. (*continued*)

Type	Case examples	Root cause	Reasons generalized
Physical duplication	Two winners announced in the same field instead of using two separate award (CA) IDs	The form allows for long text fields and it is difficult to detect whether there are one or more winners	Lining up the process and the forms is difficult – which makes any automatic detection of form errors complicated
Content error	Long text of purchase data is truncated by certain tools	No limit on size of text fields	Forms collecting data allow for lengthy textual information
	CPV codes may have several values in the field	Multiple values are allowed in one (text) field	IT is a typical form definition error where database representation (and analytical) requirements are not considered
	Purchase value is not realistic (e.g. € 1234567)	Contracting authorities have no intention to publish certain data	Deliberate misrepresentation
Inconsistency	There were calls without cancellation or eventual awards for years	Inconsistency in using or filling out the forms – as contracting authorities did not submit a cancellation notice	Forms are complex and they are difficult to change. Furthermore, due to the large volume of the data there is no bandwidth (process or technology) to detect inconsistencies
	Type of call value does not match the actual form used (or should have used a different form)	Each type of call notice has its own form (#) – but often the number entered into the form is wrong due to human error	Although there are human errors, often certain basic errors may or may not be detected by the sanity algorithms

Overall, it can be concluded from the table (and the examples presented in the analysis section), that in the context of the public sector quality of open data is far from being a simple technical issue (or a DB problem). Issues with the official forms mandated, the complexity of the process, or even deliberate misrepresentation of data may hinder the usability of the data eventually published – on top of regular technical challenges of formats, representation, and readability. In addition, the process of generating the data set version intended for open publication may bring in further errors.

6 Conclusions and Practical Results

This research paper has argued that while current OD quality frameworks are strong tools when it comes to assessing OD quality or measuring maturity of data released, they are inadequate when it comes to helping public organizations how to release their data in better shape and how to improve quality as experienced by the end user during reuse. It was proposed, that an investigation of the root causes leading to lower quality PSI/OGD is needed and the idea and its possibilities were demonstrated through a case study.

It was demonstrated through the causes identified that ensuring quality of open government data is not simply a technical exercise and often even good organizational practices might not be enough. Although proper data governance principles augmented with well-organized data management and release processes could certainly improve, quality starts at the forms and rules set out in regulations. Therefore, for deeper quality improvements changes need to reach as far as the level of policy frameworks.

During the execution of the case study data quality issues identified in the TED csv datasets had been communicated to the issuer of the data. As a result, first the aforementioned "advanced notes" [27] had been released, and later improvements have been made to the production of the TED OD – with latest datasets released during the completion of this paper (i.e. changes could not be included here). In addition, the codebook [26] has been updated as well. An obvious next step is to investigate the content and changes of the new datasets. The publishing team could also be contacted again in order to collect information about the actions taken to improve quality: this could help further validating the root cause analysis presented here, potentially leading to advanced guidelines for issuers of PSI/OGD.

References

1. Batini, C., Cappiello, C., Francalanci, C., Maurino, A.: Methodologies for data quality assessment and improvement. ACM Comput. Surv. **41**(3), 16–52 (2009)
2. Berners-Lee, T.: Linked Data - Design Issues. (2010) (originally posted 2006). https://www.w3.org/DesignIssues/LinkedData.html. Accessed 10 May 2018
3. Bertot, J.C., Gorham, U., Jaeger, P.T., Sarin, L.C., Choi, H.: Big data, open government and e-government. Inf. Polity **19**(1–2), 5–16 (2014)
4. Bovens, M., Goodin, R.E., Schillemans, T. (eds.): The Oxford Handbook of Public Accountability. Oxford University Press, Oxford (2014)
5. Davies, T.: Open data barometer. Global Report 2013. http://www.cocoaconnect.org/publication/open-data-barometer-2013-global-report. Accessed 21 2017
6. 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). https://doi.org/10.1007/978-3-642-14799-9_5
7. Dedeke, A.: A conceptual framework for developing quality measures for information systems. In: Proceedings of the 5th International Conference on Information Quality, pp. 126–128 (2000)

8. Eisenhardt, K.M., Graebner, M.E.: Theory building from cases: opportunities and challenges. Acad. Manag. J. **50**(1), 25–32 (2007)
9. Erridge, A., Hennigan, S.: Public procurement and social policy in Northern Ireland: the unemployment pilot project. In: Thai, K.V., Piga, G. (eds.) Advancing Public Procurement: Practices, pp. 280–303. PrAcademics Press, Boca Raton (2007)
10. Erickson, J.S., Viswanathan, A., Shinavier, J., Shi, Y., Hendler, J.A.: Open government data: a data analytics approach. IEEE Intell. Syst. **28**(5), 19–23 (2013)
11. Frank, M., Walker, J.: User centred methods for measuring the quality of open data. J. Commun. Inf. **12**(2), 47–68 (2016)
12. Grönlund, Å., Horan, T.A.: Introducing e-gov: history, definitions, and issues. Commun. AIS **15**(1), 39 (2005)
13. Huijboom, N., Van den Broek, T.: Open data: an international comparison of strategies. Eur. J. ePract. **12**(1), 4–16 (2011)
14. Jetzek, T., Avital, M., Bjorn-Andersen, N.: Data-driven innovation through open government data. J. Theor. Appl. Electron. Commer. Res. **9**(2), 100–120 (2014)
15. Kahn, B.K., Strong, D.M., Wang, R.Y.: A model for delivering quality information as product and services. In: Proceedings of the 1997 Conference on Information Quality, Cambridge, MA, pp. 80–94 (1997)
16. Levitin, A., Redman, T.: Quality dimensions of a conceptual view. Inf. Process. Manag. **31** (1), 81–88 (1995)
17. Naumann, F., Rolker, C.: Assessment methods for information quality criteria. In: Proceedings of the 5th International Conference on Information Quality, Humboldt-Universität zu Berlin, Institut für Informatik, pp. 148–162 (2000)
18. OKF - Open Knowledge Foundation: Open knowledge definition (2006). http://www.opendefinition.org/. Accessed 17 Sept 2017
19. Omidyar Network: How open data can help achieve the G20 growth target (2014). www.omidyar.com/sites/default/files/file_archive/insights/ON/Report_061114_FNL.pdf. Accessed 7 Jan 2017
20. Parks, W.: The open government principle: applying the right to know under the constitution. George Wash. Law Rev. **26**(1), 1–22 (1957)
21. Pignotti, E., Corsar, D., Edwards, P.: Provenance principles for open data. In: 2011 Proceedings of Digital Engagement (2011)
22. Scannapieco, M., Catarci, T.: Data quality under a computer science perspective. Arch. Comput. **2**, 1–15 (2002)
23. Self, P.: Government by the Market? Macmillan, London (1993)
24. Stake, R.E.: Qualitative case studies. In: Denzin, N.K., Lincoln, Y.S. (eds.) Strategies of Qualitative Inquiry, pp. 119–149. Sage, Los Angeles (2008)
25. Tayi, G.K., Ballou, D.P.: Examining data quality. Commun. ACM **41**(2), 54–57 (1998)
26. TED: TED Processed Database: Notes & Codebook, Version 2.2. http://data.europa.eu/euodp/repository/ec/dg-grow/mapps/TED(csv)_data_information.doc. Accessed 17 May 2018
27. TED: TED Advanced Notes. Version 0.9. http://data.europa.eu/euodp/repository/ec/dg-grow/mapps/TED_advanced_notes.docx. Accessed 20 Sept 2017
28. Tygel, A., Kirsch, R.: Contributions of Paulo Freire for a critical data literacy. In: Proceedings of Web Science 2015 Workshop on Data Literacy (2015)
29. van Zeist, R., Hendriks, P.: Specifying software quality with the extended ISO model. Softw. Qual. J. **5**(4), 273–284 (1996)
30. Vetrò, A., Canova, L., Torchiano, M., Minotas, C.O., Iemma, R., Morando, F.: Open data quality measurement framework: definition and application to open government data. Gov. Inf. Q. **33**(2), 325–337 (2016)

31. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (1996)
32. Wormell, I.: Information quality: definitions and dimensions. In: Proceedings of a NORDINFO Seminar, Royal School of Librarianship, Copenhagen, Taylor, London (1990)
33. Yin, R.K.: Case Study Research: Design and Methods. Sage, Thousand Oaks (2003)
34. Yu, H., Robinson, D.G.: The new ambiguity of open government. UCLA Law Rev. Discourse **59**, 178 (2011)
35. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: a survey. Semant. Web **7**(1), 63–93 (2016)
36. 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. PAIT, vol. 4, pp. 115–135. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-9563-5_8
37. Zuiderwijk, A., Janssen, M.: The negative effects of open government data - investigating the dark side of open data. In: Proceedings of the 15th Annual International Conference on Digital Government Research, pp. 147–152. ACM (2014)

Data Makes the Public Sector Go Round

Paraskevi Christodoulou[1], Stefan Decker[2], Aikaterini-Vasiliki Douka[1], Charalampia Komopoulou[1], Vasilios Peristeras[1(✉)], Sofia Sgagia[1], Vaios Tsarapatsanis[1], and Dimosthenis Vardouniotis[1]

[1] International Hellenic University, 14th km Thessaloniki-Moudania, 57001 Thermi, Greece
vivichristod@gmail.com, katerina.a.douka@gmail.com, charakomop@gmail.com, v.peristeras@ihu.edu.gr, sgagia.sofia@gmail.com, tsarvakis@gmail.com, vardouniotisdimos@gmail.com

[2] Fraunhofer/FIT, RWTH Aachen, Aachen, Germany
stefan.decker@fit.fraunhofer.de

Abstract. This article summarizes the results of a systematic literature review on how the new data technologies affect the public sector and what their impact on governments are. The opportunities and the challenges that public administrations face nowadays in a data-driven world are important. Our research retrieved opportunities related to transparency, innovation, public participation and efficiency, while challenges regarding privacy issues, technical difficulties, data management, cultural and political obstacles were found. By overcoming the challenges and empowering the opportunities for data usability in public sector, exploitation of data could become the foundation for innovation and public sector transformation.

Keywords: Data · Public sector · Public administration · Literature review

1 Introduction

Nowadays, data thrive of importance. In a digital and fast changing environment, huge companies such as Wal-Mart, a retail giant, logs more than 1 million customer transactions every hour, feeding databases estimated at more than 2.5 petabytes, which is equivalent of 167 times the books in America's Library of Congress. Additionally, social media platforms, such as Facebook are overwhelmed with a total storage that has exceeded 40 billion photos [28]. These examples highlight that organizations become increasingly interested in the volume and value of data they can capture, in the creation of new data and in search for external data sources.

Due to the massive attention that data is gaining, public sector slowly but steadily takes decisive actions towards adopting and using similar practices used in the private sector. Governments and public sector, attempt to exploit data at their benefit, in order to achieve improved quality services for all citizens. As an example, through use, reuse and free distribution of datasets, governments promote entrepreneurship, innovation, and citizen-centric services [23].

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 221–232, 2018.
https://doi.org/10.1007/978-3-319-98690-6_19

The structure of the paper is described below. In Sect. 2, the research method is presented with all the steps that have been executed to decide the final literature. In Sects. 3 and 4, the findings of the research are described in detail. Specifically, opportunities and challenges are gathered together and categorized into wider groups. Last, some conclusions appear in Sect. 5.

2 Research Procedure

This paper constitutes a systematic literature review [25]. The research was initiated by taking into account the guidelines proposed in [11] with explicit inclusion and exclusion criteria. Specifically, the process we followed for preparing this paper used the following steps: first the definition of research questions and relevant keywords, then the identification of digital libraries and lastly, the selection of studies, which are following the framework of inclusion/exclusion criteria.

2.1 Definition of Research Questions

The aim of this systematic literature review is to investigate the role of emerging data technologies and applications, such as big data technologies and cloud computing [3] in the public sector while identifying potential opportunities and risks deriving from different usages of data in governments, worldwide. Our work addresses the following main Research Question which further includes two defining Questions:

What is the impact of the implementation of Data Technologies in public sector entities?

1. Which opportunities arise for data-driven administrations?
2. What are the challenges of a data-driven public sector?

2.2 Search Methodology

The aim was to include in our research scientific articles, other literature review, reports and studies from credible sources. As a result, the research was conducted using the following digital libraries: ACM, IEEE Xplore, Science Direct, Springer Link, JSTOR, EBSCO Information Services. The search keywords used for this literature review were the following:

3. "data-driven public sector"
4. "public sector AND (machine learning OR artificial intelligence OR data analytics OR big data OR Internet of Things)"
5. "data driven AND (public service OR (public sector AND (challenges OR machine learning)))".

As it could be concluded by the selection of the keywords, our team used extensively the correlative conjunction of "AND" and "OR" in order to link the main research on public sector with different fields closely related to the concept of data-

driven administrations. Moreover, we extended the search by using synonyms for the term "public sector", i.e. "public administration" and "public service".

2.3 Study Selection

After using the above method, a list of relevant to the study articles was created. However, some of the results were still irrelevant to the research questions even with the research keys being in the abstract or/and title. So, within this procedure, some inclusion and exclusion criteria were developed and applied. The inclusion criteria according to which publications were included in the list of the essential studies are the following:

1. Articles and conference papers focusing on data driven organizations, data usage and innovation deriving from the data usage in public sector.
2. Studies that describe how data driven societies and organizations collaborate in promoting innovation.

The exclusion criteria were the following:

1. Studies that partially discuss the topic and did not focus on the main aspects of data driven public sector.
2. Studies that focused on describing data driven administrations but not discussing opportunities or risks.
3. Studies that include only technical aspects (e.g. data technologies, architectures).

The papers used in our research were gathered in December 2017. So this literature review includes information on studies that were published before that date. Articles and studies published before 2000 were excluded from the research. Also, all articles in languages other than English were excluded. As shown in Fig. 1, in the first step of research done by using the keywords, the number of results retrieved was 1030. In the next step, duplicates were discarded and studies that were not journal articles or conference papers were also excluded. After checking the title, the results reduced to 220. The process resumed with the thorough reading of abstracts, which led to 84 relevant articles. In the final step, after reading the full text of all articles, some of them were found irrelevant in terms of content and were excluded. So the number of articles reviewed in full text was 29.

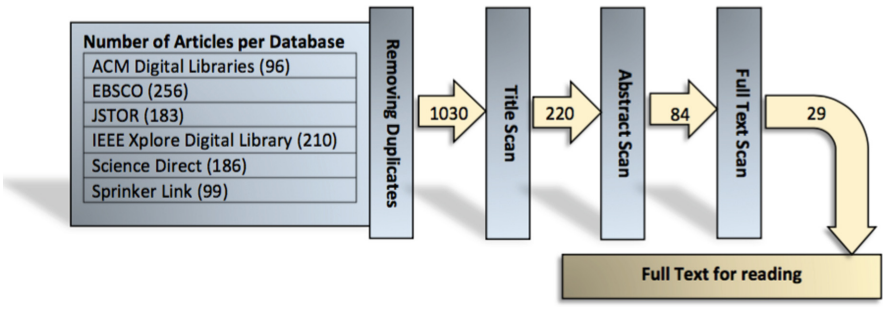

Fig. 1. Literature review summary

3 Opportunities for a Data Driven Public Sector

The adoption of new data technologies entails many opportunities for governments. The findings presented in this part of the review were grouped, as shown in Table 1, where only 22 out of 29 articles included the discussed opportunities.

Table 1. Papers discussing opportunities for data-driven administrations

Opportunities	Innovation	Transparency	Efficiency
[1]	✓	✓	✓
[2]	✓	✓	✓
[3]	✓		
[4]	✓	✓	✓
[5]			✓
[6]	✓		✓
[7]		✓	✓
[8]			✓
[9]		✓	✓
[10]			✓
[12]		✓	✓
[13]	✓		✓
[14]	✓	✓	✓
[15]	✓	✓	✓
[17]	✓	✓	✓
[20]	✓	✓	✓
[27]			✓
[29]			✓
[31]			✓
[32]		✓	
[33]	✓		

3.1 Efficiency

According to references found in the articles, in almost every case, efficiency level is positively affected by the data revolution. Analytical implementation of data owned by public sector's services could lead to the improvement of governance practices in terms of control of expenditure and policy-making [14]. In Ireland, scientists applied data mining techniques to data deriving from social media platforms in order to predict early variations in unemployment rates. In Indonesia, data from Twitter was used for the prediction of the inflation rates by analyzing real-time tweets regarding the price of rice [10, 24]. In addition, in New York, data scientists managed to detect illegal conversions, a condition where the number of people that live in a building surpasses the appointed number of tenants [9].

Detailed, timely and linked data could be used to unravel undiscovered patterns and meaningful information [24]. Regarding this statement, big data analytics would be extremely helpful if integrated into the field of fraud detection operations, as seen in the Internal Revenue Service (IRS) of the United States. In addition, by applying clustering and classification methods on big data analytics, the public administrations could correlate and manage requests faster. Big data could enable the conversion of the existing e-government systems to multi-dynamic systems in which analytical tools such as descriptive or predictive analytics would provide valuable reports and insights [14]. Regarding the efficiency of the provided services, big data analytics assist policy-makers by providing insights about the imposed policies either by estimating the opinion of the citizens through participative technologies/applications or by prognosticating it within high levels of accuracy [9, 14, 15].

In [27], authors state that Artificial intelligence (AI) could boost the efficiency of administration. The combination of humans and AI would change decision-making processes in public services. These processes have traditionally been executed in several steps by humans until now: problem identification; information collection and analysis; searching for different solutions; assessing the alternatives; selecting from others; implementation; evaluation and feedback. As a result, currently decisions made exclusively by humans could be supplemented by AI-embedded decision-making. In the case of problem identification, for example, AI applications might detect problems that managers would not pay attention to [27].

3.2 Public Participation and Transparency

A data-oriented public sector could offer opportunities for public participation and transparency. All these could be achieved through big data technologies and tools, as this field has the potential of delivering reliable services in the e-government department and openness of a variety of datasets [1]. Open data published by governmental portals along with ease of access would lead to higher transparency for citizens [7]. Furthermore, one of the most notable advantages of open data is that by making data transparent, public trust in government and civil servants is increased [21, 32].

Sharing data is also a great chance for governments to strengthen the sense of cooperation and communication between citizens and the state. For example, US government gathers and publishes different kind of data, which give to citizens and government the opportunity to utilize them more efficiently [2]. Also, through the social media channels, public sector transparency and accountability could be significantly improved, too. Suggestions or objections from citizens could be shared through social networks [5]. Therefore, the extended use of social media in e-government has the potential of gradually improving transparency [14].

3.3 Innovation

Innovation in a data driven government environment is centered around collaboration between governments, private and public organizations in order to invent new applications and solutions. These innovations could lead to the achievement of higher levels of efficiency, transparency, accountability, service quality and trust in government.

Data-driven innovation requires technological capabilities needed for collecting, opening, sharing, combining, and analyzing data. Hence, is a complex field in which many stakeholders are involved as well as influential factors at a strategic, political, organizational, technical and governance level [13].

A disruptive aspect of innovation would be conceived through cloud services. The utilization of cloud computing would be a formidable asset for both public and business administrations, as well as for the citizens themselves. For the public sector, it is an important tool that could facilitate data interoperability, security and flexibility at the same time. Overall, cloud computing could advance the software services of public administrations and accelerate innovation processes [3].

Several case studies support the promotion of data-driven innovation in the public sector. For example, the "Living Lab" project is an environment that supports public open innovation processes by enabling public administrations to digitally communicate with private sector organizations and derive jointly creative solutions [13]. An experiment conducted in the United Kingdom showed that the increase in the proportion of distributed data between public administrations was a worth trying venture. If local authorities share their data, developers would be able to produce innovative applications (e.g. Transport of London) that could benefit both citizens and the council [31].

Smart cities are yet another innovative regional government approach, as they create a new path in understanding the urban problems [4]. Cities are becoming more complex than ever due to new technologies and digitalization. Hence, there is a global willingness to enhance the understanding of modern cities problems by making thoughtful decisions and taking actions towards the right direction. Concurrently, critical problems relating to cities like transportation and energy management could be resolved by utilizing data technologies. The approach of Smart Cities is highly correlated with big data technologies [1]. Some examples of Big Data applications are "Smart Education", "Smart Traffic Lights" and "Smart Grid" [2].

E-participation is also fundamental for the Smart Cities Initiative. It promotes the co-operation of all the community members [20]. Smart cities take into account the exploitation of data technologies for communication and dissemination purposes. Citizens have the capability of feedback with online participation on urban problems, new plans and policies [4]. Additionally, open data produces social benefits, as citizens can more easily interact with government in an informative manner [32]. As a result, the citizen views and expectations can be identified, adopted and covered in future policies.

For all these features, data technologies and tools represent an unprecedented innovation, which has already been introduced to many countries such as Brazil, Singapore, and Portugal [20].

4 Challenges

Nowadays, data-driven public sector comes with several types of challenges including data security and privacy, portability and interoperability. Additionally, there are legal problems, such as national domination, old-fashioned legacy laws and heavy

procurement processes. In addition, citizens and businesses have high expectations concerning better, more effective and personalized services [3]. The aforementioned issues have been identified in 26 papers listed in Table 2.

Table 2. Papers discussing challenges for data-driven administrations.

Challenges	Cultural and political barriers	Technical obstacles	Privacy and security issues	Efficient data management
[1]	✓	✓	✓	
[2]	✓	✓	✓	✓
[3]		✓		
[4]			✓	✓
[5]			✓	
[6]	✓			
[7]			✓	
[8]	✓			
[9]	✓	✓	✓	✓
[10]		✓		
[12]	✓	✓	✓	
[14]	✓	✓	✓	
[16]		✓		
[17]			✓	
[18]				✓
[19]	✓			✓
[20]	✓	✓	✓	
[21]			✓	
[22]	✓			
[24]			✓	
[26]			✓	
[27]	✓			
[29]		✓	✓	✓
[30]			✓	✓
[32]		✓		
[33]		✓		

4.1 Cultural and Political Barriers

Despite the technology invasion and its integration in people's everyday life, public governance confronts serious challenges regarding the wide pursuance of data technologies. Governance culture, politics, and ethics play a significant role in multiple ways.

An important challenge concerning public sector's general approach to data is to identify and use it as an asset in order to strengthen internal procedures [8]. As the digital world evolves, the governmental open data become part of policy agendas.

However, there is reluctance in releasing public information because of the preference of maintaining secrecy in that field. Further, fears exist concerning quality, accuracy and exposure to mistakes that may occur by the employees, as well as the misunderstanding of the real value created from open data [12].

Traditionally, the public sector has been a centralized and bureaucracy-oriented organization, thus the resistance to change exist in many levels of governance. The existing organizational models may impede the implementation of new technologies [14]. For instance, the utilization of semantic web technologies in the public sector is hindered by the complex public organizational structure as well as limited resources [22].

Cultural impediments are affecting public sector's workforce management. Provision for training of employees and administrators on issues related to data is out of the governmental agenda [22]. Furthermore, data may not be transformed into knowledge in the hands of not experienced people. Misconceptions and misinterpretations could affect negatively the policy-decision making as analysis could be processed by uninformed individuals, thus making the role of experts crucial [1, 19]. Shortage of leadership may also be observed [12].

A major factor which subtends to the cultural diversifications is the level of public engagement in innovative public sector initiatives. A large proportion of the public is not digitally engaged [6]. This lack of familiarity strengthens citizens' resistance to change, which consequently may hinder the holistic implementation of programs like Smart Cities [2, 20].

Lastly, ethical concerns arise due to possible insertion of artificial intelligence devices in public sector. 'Hubo' - 'Hubogent' is a compound of 'Hubo' and 'agent'- a humanoid robot developed by the Korean Advanced Institute of Science and Technology (KAIST) in South Korea, a robotic agent geared with an intelligence system that carries out administrative work for human beings. According to the authors, there is an inherent danger in creating devices whose intelligence could, in certain areas, surpass that of humans. In order to prevent potential misuse and an imbalance of responsibility, appropriate prevention measures -including monitoring and controlling AI, effective legal gadgets and ethical systems, and principles of liability- must be established [27].

4.2 Technical Barriers

Another ominous challenge related to data implementations in public sector is that huge amount of data is collected from various sources, such as mobile devices and sensors and in many different formats. The existing tools do not align with the capabilities needed to process and store the amount of data generated and the traditional systems become inadequate and insufficient regarding the emerging data processing needs [1, 14]. Hence, the necessity for upgraded methods and tools appears.

Another major issue is that the paper-based media in public organizations inhibits information sharing as reproduction and dissemination of information are relatively expensive and time consuming. Movement of information is slow and cumbersome, inducing the generation of information silos and barriers. Using digitized applications to save and share information would demolish the above obstacles [33].

Low quality is a huge drawback due to the fact that data exists in many different formats, though unlinked and stored in separate systems with no provision for integrative features [10]. Due to the complex structure of public sector and the absence of collaborative design, large data sets are held into non-interoperable systems, which results in making the cost of integration for future analysis huge to bear [3, 9].

Governments dealing with issues of big data integration face some unique challenges. The biggest one appears to be the data collection. Governments encounter difficulties, as the data not only derives from multiple channels (such as social networks, Web, and crowdsourcing) but also from different sources (such as countries, institutions, agencies, and departments) [20]. Also, sharing data and information between countries remains a perpetual challenge. This issue complicates the integration effort of complementary data among government agencies and departments [16].

Moreover, integration of open data is challenging and often requires detailed analytical skills. Additionally, difficulties in accessing open data and failure to update them on a regular basis, prevent organizations and people from relying on public sector's open data. Presently, open data are mostly available in different formats and have interfaces that are not user-friendly, a situation which fails to attract many users [32].

4.3 Data Protection – Privacy and Security Issues

There are also major issues regarding data privacy and security in public sectors. Not only all the collected data have to be into a machine-readable format, but also actions should be taken in order to protect the personal data or employee details [17]. Citizens anticipate that their personal data would be collected, protected and appropriately used by public agencies [29]. Unauthorized access of personal information may cause numerous problems. First of all, it may damage the commonly held opinion or worst, cause physical corruptions. In addition, it may lead to lack of trust in employees and instant rejections of people. There might also be some negative issues in operational systems, like delays and inadequacies [29].

Every single department of a public sector owns not only public but also private data, so high-security levels must be maintained in order to avoid unauthorized access. Public agencies should collect, share and use data in a way that citizens' personal data are not violated. This is of utmost importance because most public organizations encounter cyber attackers. Undoubtedly this fact affects the productivity and the professional reputation of organizations [21].

Furthermore, similar aspects of security obstacles are raised regarding big data. Specifically, collecting and using big data, while protecting privacy, is a major challenge [24]. Governments have access to huge volume of data and in several cases, the information provided is not filtered and derives from non-credible sources [30].

In addition, variances arise between U.S legislation and the potential use of data from public organizations as the U.S public sector needs new data protection technology in order to ensure privacy. This is a strong challenge considering that there is certain unwillingness from citizens to cede over data to the state [26].

4.4 Efficient Data Management

Difficulties regarding efficient Data Management is another barrier for a data driven public sector. A remarkable problem is that the majority of the data remains unstructured and thus renders unusable. Such data includes photos, videos, audio files, etc. that must be converted into structured forms to be analyzed and utilized appropriately [2].

Strong evidence of the above is an Australian case study about a Firearms Management System (FMS), under the responsibility of the Western Australia Police. FMS is a system that holds information about all citizens of Western Australia that are certified to own a firearm. The study revealed major problems on entering, processing and presenting data. As a result, the local authorities were unable to manage the information regarding data of the licenses [29].

Goal displacement appears to be one of the main concerns. Specifically, this problem occurs when the administrative level of the organization focuses on tasks that are oriented towards measured outcomes that involve a massive volume of data [19].

Despite the fact that intelligent data technologies are already available to governments, there is still a challenge for administrators to correctly identify areas of potential use. In the case that they already possess data, they should manage them in the most beneficial way. According to a public sector CIO [9], only a 30% percent of the current data is going into analytics procedure, leaving a vast amount of information untapped [18]. Nevertheless, the challenge that arises here concerns the interpretation of the data because each interested party may hold a different perspective regarding the same information [30].

5 Conclusion

During our research, we came across many successful cases and examples which demonstrate the net value that could be added to public sector entities as a result of the successful management and exploitation of existing data. That value could be measured in terms of the delivered services quality, functionality, fiscal planning, local governance and new practices regarding government crowdsourcing. Nevertheless, several challenges remain due to the introduction of new processes and new issues related to privacy, applicable legal framework, security, and cultural issues. Although considerable insight has been gained upon those matters, our review suggests that further primary research needs to be performed especially to assess the user experience, the impact of new data technologies, the challenge of rationalization and simplification of public sector process design as well as the appropriate cultural issues and mentality of the public sector as well. Policy makers should encourage, amongst others, the (re-) education of public sector employees and create training programs in accordance with public needs.

References

1. Al-Sai, Z.A., Abualigah, L.M.: Big data and e-government: a review. In: 2017 8th International Conference on Information Technology (ICIT), pp. 580–587. IEEE (2017)
2. Al Nuaimi, E., Al Neyadi, H., Mohamed, N., Al-Jaroodi, J.: Applications of big data to smart cities. J. Internet Serv. Appl. **6**, 25 (2015)
3. Alonso, J., Escalante, M., Orue-Echevarria, L.: Transformational cloud government (TCG): transforming public administrations with a cloud of public services. Procedia Comput. Sci. **97**, 43–52 (2016)
4. Batty, M., et al.: Smart cities of the future. Eur. Phys. J. Spec. Top. **214**, 481–518 (2012)
5. Bovalis, K., Peristera, V., Abecasis, M.: Promoting interoperability in Europe's e-government. Computer **47**, 25–33 (2014). Long. Beach. Calif
6. Chatbots lead the debate on public sector use of AI to streamline digital s…: EBSCOhost. http://web.b.ebscohost.com/ehost/detail/detail?vid=0&sid=8e31f9a2-753f-40fa-a949-a1c7eb1b9d5c%40sessionmgr101&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3D%3D#AN=123966934&db=bth. Accessed 14 Nov 2017
7. Conradie, P., Choenni, S.: On the barriers for local government releasing open data. Gov. Inf. Q. **31**, S10–S17 (2014)
8. Data rich, insight poor: EBSCOhost. http://web.b.ebscohost.com/ehost/detail/detail?vid=0&sid=12183bac-843c-45ee-a601-c1d1970489c2%40sessionmgr120&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3D%3D#AN=96063575&db=bth. Accessed 14 Nov 2017
9. Desouza, K.C., Jacob, B.: Big data in the public sector: lessons for practitioners and scholars. Adm. Soc. **49**, 1043–1064 (2017)
10. Gamage, P.: New development: leveraging 'big data' analytics in the public sector. Public Money Manag. **36**, 385–390 (2016)
11. Guidelines for performing Systematic Literature Reviews in Software Engineering, version 2.3, Software Engineering Group, School of Computer Science and Mathematics, Keele University and Department of Computer Science, University of Durham, Durham (2007)
12. Hardy, K., Maurushat, A.: Opening up government data for big data analysis and public benefit. Comput. Law Secur. Rev. **33**, 30–37 (2017)
13. Janssen, M., Konopnicki, D., Snowdon, J.L., Ojo, A.: Driving public sector innovation using big and open linked data (BOLD). Inf. Syst. Front. **19**, 189–195 (2017)
14. Joseph, R.C., Johnson, N.A.: Big data and transformational government. IT Prof. **15**, 43–48 (2013)
15. Kacem, A., Belkaroui, R., Jemal, D., Ghorbel, H., Faiz, R., Abid, I.: Towards improving e-government services using social media-based citizen's profile investigation. In: Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance - ICEGOV 2015–2016, pp. 187–190. ACM Press, New York (2016)
16. Kim, G.-H., Trimi, S., Chung, J.-H.: Big-data applications in the government sector. Commun. ACM **57**, 78–85 (2014)
17. Krishnamurthy, R., Awazu, Y.: Liberating data for public value: the case of Data.gov. Int. J. Inf. Manag. **36**, 668–672 (2016)
18. Kuraeva, A., Kazantsev, N.: Survey on big data analytics in public sector of Russian federation. Procedia Comput. Sci. **55**, 905–911 (2015)
19. Lavertu, S.: We all need help: 'big data' and the mismeasure of public administration. Public Adm. Rev. **76**, 864–872 (2016)
20. Lopes, N.V.: Smart governance: a key factor for smart cities implementation. In: 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), pp. 277–282. IEEE (2017)

21. MacAs, M., Lagla, L., Fuertes, W., Guerrero, G., Toulkeridis, T.: Data mining model in the discovery of trends and patterns of intruder attacks on the data network as a public-sector innovation. In: 2017 4th International Conference on eDemocracy and eGovernment, ICEDEG 2017 (2017)
22. Mouzakitis, S., et al.: Challenges and opportunities in renovating public sector information by enabling linked data and analytics. Inf. Syst. Front. **19**, 321–336 (2017)
23. O' Toole, G.: Why public sector organizations must 'Brake the Boundaries' in 2018. https://www.openaccessgovernment.org/why-public-sector-organisations-must-break-the-boundaries-in-2018/40819/. Accessed 12 Feb 2018
24. Shi, J., Ai, X., Cao, Z.: Can big data improve public policy analysis? In: Proceedings of the 18th Annual International Conference on Digital Government Research - dg.o 2017, pp. 552–561. ACM Press, New York (2017)
25. Siddaway, A.: What is a systematic literature review and how do I do one? (2014)
26. Stough, R., McBride, D.: Big data and U.S. public policy. Rev. Policy Res. **31**, 339–342 (2014)
27. Suk Kim, P., Hong, K.P.: Debate: humanized robotic agents in government—the emergence of the 'Hubogent'. Public Money Manag. **37**, 131–132 (2017)
28. The Economist. Data, data everywhere. https://www.economist.com/node/15557443. Accessed 12 Feb 2018
29. Thompson, N., Ravindran, R., Nicosia, S.: Government data does not mean data governance: lessons learned from a public sector application audit. Gov. Inf. Q. **32**, 316–322 (2015)
30. Tsoukias, A., Montibeller, G., Lucertini, G., Belton, V.: Policy analytics: an agenda for research and practice. EURO J. Decis. Process. **1**, 115–134 (2013)
31. Using public sector open data to benefit local communities: EBSCOhos. http://web.b.ebscohost.com/ehost/detail/detail?vid=0&sid=fbadf6cb-1152-4d42-9e20-ff15e6b991e4%40sessionmgr101&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3D%3D#AN=94983206&db=bth. Accessed 14 Nov 2017
32. Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., Dwivedi, Y.K.: Open data and its usability: an empirical view from the Citizen's perspective. Inf. Syst. Front. **19**, 285–300 (2017)
33. Wiredu, G.O.: An institutional perspective on the challenges of information systems innovation in public organisations. In: Wimmer, M.A., Chappelet, J.-L., Janssen, M., Scholl, H.J. (eds.) EGOV 2010. LNCS, vol. 6228, pp. 97–108. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14799-9_9

Smart Governance (Government, Cities and Regions)

Fostering the Citizen Participation Models for Public Value Creation in Cooperative Environment of Smart Cities

Manuel Pedro Rodríguez Bolívar(✉)

University of Granada, 18071 Granada, Spain
manuelp@ugr.es

Abstract. The growth of smart cities is forcing governments to focus their efforts on the increase of public value creation. Based on prior research, on the perception of smart city practitioners and on an empirical observation, this paper seeks to analyse the public value creation under the context of the smart cities, examining the model of citizen participation and the use of new technologies by city governments in smart cities with the aim at improving e-participation of the citizenry in the public arena. Findings indicate the need for using new technologies to adopt a more participative model of governance, although, nowadays, sample smart cities are not putting available technological tools or promoting citizens to be involved in online public consultations, discussions and petitions.

Keywords: Public value creation · Creative citizenship · Smart governance
Smart cities

1 Introduction

Public value creation has become a main domain of the smart city (SC) phenomenon [15] and introduces a more holistic approach to city governance, with the implementation of new and innovative forms of city governance based on the concept of network governance and pursuing the achievement of citizen-centric projects [16]. The smart governance is not therefore built on what governments do but about the outcomes of interactions between all actors in the public domain seeking for a common purpose and strategic priorities, in brief, in the pursuit of public value goals [19].

Civic participation is thus seen as a main way of transforming government to make it open and closer to the citizenry needs under the context of smart cities and the smart cities movement has promoted the creation of places where relational networks of actors are produced, constituting the urban governance as a crucial axis of smarter cities, since they reflect how public value can be generated with the participation of citizens and other social actors [11] and highlights the process of coordinating and steering the urban society toward collectively defined goals [17]. Nonetheless, up to now, there is a lack of research regarding the analysis of the relevance that smart cities associate to collaborative governance models as well as whether smart cities are using e-participation tools to promote the citizen involvement in the public affairs of the city.

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 235–248, 2018.
https://doi.org/10.1007/978-3-319-98690-6_20

This paper contributes to the advance of the knowledge of public value creation analysing the model of citizen participation in smart cities and the use of tools for improving e-participation in these cities. To achieve this aim, this paper analyses a sample group of smart cities that are members of the EUROCITIES network and seeks to examine the model of citizen participation and to find whether perceptions about smart governance models (e-survey) are coming true into real political intentions in smart cities through the offer of e-participation tools for citizen involvement.

The remainder of the paper is as follows. Next section describes the public value creation in smart cities based on building participative governance of the cities and after that, an empirical research on sample cities that are members of the EUROCITIES network is performed, especially in those that promote cooperation engaging and empowering citizens to help make their cities 'smarter' [5]. Then, the data and method used in this research are described and each one of the research questions proposed in this paper is analysed through the lens of the results of our study. Finally, the conclusions and discussions section brings this paper to an end.

2 Fostering Public Value Creation in Smart Cities Through Cooperative Environments

A growing urban development and reforms and innovations in new technologies have produce challenges of innovation in city governance [24] with the aim at facing new challenges of the current society. These new urban structures have been based on the use of non-hierarchical, non-market forms of organization in the public sector [3], built over an intensive use of new technologies. The availability of these ubiquitous ICT infrastructures has stimulated the advance on new citizen-centric services [12] and the creation of cooperative environments to achieve problem-solving and innovation processes [18], creating an integrated approach of the smart cities where governance among stakeholders is converted into the main cornerstone of the smart cities [10]. It has brought the creation of public value into a lively debate in the current political and public sector management arena in smart cities.

Under this framework, European Commission [25] has recognised that by mobilising citizens at local level to debate on concrete issues from the European political agenda will promote civic participation in the EU policy making process and develop opportunities for societal engagement. Indeed, there are no 'smart cities' without 'smart citizens', adding challenges around engaging and empowering citizens to help make their cities 'smarter' [5]. In this regard, under cooperative environments of smart cities [14], cities engage different kinds of knowledge, and encourage widespread public participation to deal imaginatively with complex issues and offers direct association with technological innovation and communal interaction [13]. Thus, city governments should promote citizen involvement in policy making as well as in the implementation of innovative processes and instruments designed to enhance governance and public value creation in European cities, especially under the smart cities' framework [9].

Nonetheless, prior research has concluded that city governments far from using new technologies to promote the participation of citizens in public affairs, usually use them as other means of communication, mainly linked to promote cultural events,

tourism in the city or, even, as a city branding strategy –representation strategy- [20]. In addition, although the use of participatory innovations has expanded vastly around the world, these new channels for participation have become popular as add-ons to established channels and have been mainly focused on elections and participatory budgeting [26]. Therefore, the following research question is derived:

RQ1. Have governments in smart cities that are involved in projects about cooperative environment of smart cities put available e-participation tools for promoting the collaborative or participative models of governance as a way of public value creation?

According to prior research [19] smart practitioners think that city governments should act as leaders in innovation policies and actions. Nonetheless, the smart city is a city that emerges like an innovation ecosystem [4], where intensive civic participation is an essential ingredient of radical and sustainable innovation [8].

Acting with the creation of government transparency and open data websites, and making data and information available, city authorities support crowd-creation and production of ideas and solutions. It fosters the citizen involvement in public decisions guaranteeing an open and egalitarian approach among stakeholders. Therefore, it is expected to find an association between those smart cities that are prone to stakeholder involvement (selected or open stakeholder participation) and those that have created government transparency and open government data websites to disclose information for reaching out to the civil society in order to base their decisions on this information. So, the following research question is derived:

RQ2. Is the information transparency of governments through the creation of transparency and open government data websites associated with those cooperative environments of smart cities that are more prone to governance models based on selected or open stakeholder participation?

Finally, in cooperative environments of smart cities, intensive civic participation is an essential ingredient of radical and sustainable innovation [8]. To achieve this aim, smart cities are creating collaborative environments using tools like the creation of e-participation platforms or putting available tools for contacting with the government, in order to involve citizenry in public consultations, discussions and petitions. Therefore, the last research question of this research is:

RQ3. Is the active participation through the creation of participation platforms or tools for contacting with the government associated with those cooperative environments of smart cities that are more prone to governance models based on selected or open stakeholder participation?

3 Data and Method

3.1 Data Collection

The data of this paper was gathered using two different data collection methods. The first one is based on a questionnaire sent to all representatives of local governments labelled as "smart cities" that are members of the EUROCITIES network to capture their perceptions about smart governance and models of participation in smart cities –see Table 1 in Appendix-. The EUROCITIES network is the network of over 180 elected local and municipal governments of largest European cities and 40 partner cities, which between them govern 130 million citizens across 35 countries (see http://www.eurocities.eu/eurocities/about_us) and is aimed at reinforcing the role of local governments in a multilevel governance structure, identifying some of the challenges and opportunities in cities that are closely linked to developments at EU level [6]. Therefore, the whole questionnaire of our main research was sent to 70 leading smart cities in Europe –those involved in smart projects into working groups- and 64 responses were received (91.42% of sample smart cities).

On another hand, this research has focused on smart cities that are labelled as full member and associate member into the working groups of "Creative citizenship" and "Integrated urban development" included in the area of cooperation. These groups are addressed to improving the city through the co-creation of solutions to local challenges using an integrated urban development, planning and governance. So, the second data collection method of this paper explores whether these smart cities have created official web pages for government transparency, open data or e-participation platforms. The total sample selection is 85 cities of 165 cities included into the area of cooperation (51.51%) or of 184 smart cities included in all areas of EUROCITIES network (46.19%).

Whereas the second collection method is addressed to answer RQ1 of this paper, joining the all data collection methods, this paper presents the results obtained for RQ2 and 3.

3.2 Method

Under the e-survey analysis, this paper presents only one question included in a wider research which designed an online questionnaire about the importance of collaborative governance in smart cities, the key pillars and outcomes of the concept of smart governance and the model of participation in developing a smart city. For the particular case of this paper, the selected question of the questionnaire was answered with Yes/No responses to each one of the items included in this question –see Table 1 in Appendix- and only one answer was allowed for each item. Also, a presentation letter was sent to explain each of the items as well as the rules of the survey and the email provided to ask for doubts about the questionnaire. The question and the presentation letter were translated into the different languages and distributed to the selected sample of smart cities. The results of this part of the questionnaire could help to characterize the participation models in sample smart cities.

Finally the second data collection method gathers information about an examination of 85 smart cities (those included in the working groups of "creative citizenship" and "integrated urban development" in the EUROCITIES network) during December of 2017 with the specific purpose of collecting data about smart technologies used for information transparency and active participation. A content analysis was conducted in each one of these websites to observe the presence, or not, of specific websites for e-participation and the issues that are dealt with into each one of the smart technologies analysed. As some websites are only performed in the official language of the city (different from English language), we have used Google translator to read them appropriately.

As for the hypothesis testing for answering RQ2 and RQ3, as all attributes are based on qualitative attributes, the Pearson Chi-Square test was chosen to check the association between responses obtained in the e-survey previously mentioned and the technological tools offered by sample smart cities to achieve government transparency and citizen participation. This methodology of research has been broadly recognized in many research studies as the best statistical tool to test hypotheses of independence for r x c contingency tables in which row and column categories are both nominal and mutually exclusive categories [27, 28]. Also, contingency tables were performed to gain an overview of the distribution of the attributes. The research questions were tested at the 0.05 significance level.

4 Result Analysis and Discussions

4.1 Collaborative or Participative Models of Governance in Smart Cities as a Way of Public Value Creation and New Technologies Used for Improving e-Participation

According to the results of our questionnaire (see Q1.1 in Table 1 in Appendix), most of the respondents indicate that the active participation of the citizens in the development of the smart city is relevant. Nonetheless the model of participation is not clear. This way, both selected and open participation models obtain the same results and no preference is shown among respondents (the 51% of the respondents indicate the need of the participation of stakeholders with a mean score of each one of the participation models: 0.48; median score: 0). It means that smart city practitioners do assign relevance for citizens to actively participate in the management of the smart city, although it seems that they advocate the collaborative models, asking for citizens their help and suggestions to take decisions in the city (see Q1.1 in Table 1 in Appendix). By contrast, the 3.12% of the respondents indicated that local governments should manage the development of smart cities without the participation of stakeholders (mean score: 0.03; median score: 0). This way, a question arises: How are local governments managing the citizen involvement in the smart cities? Which tools are they implementing or using in smart cities –Web 2.0, living labs, etc.-?

On another hand, Table 2 in Appendix shows the variables and method of evaluation used in our observation of official webpages of the selected sample of smart cities,

and Tables 3 and 4 in Appendix collects the results obtained in the examination of 85 local governments of smart cities (see Sect. 3.1.).

Results show that 54 of the 85 sample smart cities offer a transparency website or an open data platform for information transparency (63.52% of the total population in our study). Besides, 6 of the 11 smart cities that offer transparency websites also provide open data platforms. Indeed, in general, it seems clear that open data platforms are more frequent in these smart cities (49 of 85 smart cities provide it).

Regarding the information disclosed, whereas sample cities disclose organisational structure, human resources, public procurement, budgetary information, legislations and environmental information in transparency websites, the information disclosed in open data portals are usually addressed to put available information about transport/traffic, cultural activities and leisure, libraries/education, housing and environment, parking areas, job information and tourism.

On the other hand, the information format in which the information is disclosed is also relevant for information transparency [21], because if the information disclosed can be manipulated, it could help governments to encourage more civic engagement among citizenry. JavaScript Object Notation (JSON), eXtensible Markup Language (XML), Microsoft Excel or Comma Separated Values (CSV) formats are examples of generally-accepted spread-sheet formats that could help users to create their own reports. Table 3 in Appendix shows that smart cities are promoting information transparency facilitating citizens the manipulation of the information disclosed. Indeed, the main information format used is the CSV format, followed by JSON format and XML formats which offer citizens the opportunity to manipulate raw data for creating their own reports according to their needs. Also, smart cities with open data platforms usually offer the information to be downloaded in two or three information formats analysed in this paper (mainly csv and json information formats) –see Table 3 in Appendix-, and only 12 sample cities (or 6 of the respondents that are included in the previously mentioned groups of EUROCITIES network) offer only one of the information formats analysed in this paper - see Table 3 in Appendix-.

By contrast, sample smart cities are not fostering the citizen involvement in public affairs because only 38 of 85 sample smart cities (44.71%) have a specific citizen participation platform website for public consultations, discussions or petitions to the local government of the city - see Table 4 in Appendix-. Nonetheless, the most of them usually provide a specific link to contact with the government, either through web forms or through email addresses (81 of 85 for sample smart cities -95.29%-). These percentages are even inferior in the particular case of those smart cities that responded the questionnaire –see Table 4 in Appendix-.

However, it is also worthy to note that the 33.33% of sample cities with an e-participation platform allows citizens to actively participate in public consultations, discussions and petitions, although public consultation is the most offered option in these web portals (31.76%). This situation is similar among the cities that responded the questionnaire (only 32 smart cities are those included in the previously mentioned groups of EUROCITIES network) but they are also prone to offer citizens to make petitions. Also, a great deal of smart cities that offer a citizen participation platform website also provide the possibility to contact with the local government. So, it seems

that the contact with the government is a possibility that is always offered by those smart cities that foster active citizen participation in the municipality.

In resume, results indicate that smart cities are promoting information transparency [22] because they are disclosing a great deal of information with the use of easily manipulated information formats but they are failing in facilitating the active participation of citizens nowadays.

4.2 Selected or Open Stakeholder Participation Models of Governance as Indicators for Offering Information Transparency and Active Participation

Table 5 in Appendix shows the contingency tables between each one of the technological tools used for information transparency or active participation, and the responses that representatives of local governments have pointed out regarding the governance model to be implemented in their smart city.

As for information transparency, Table 5 shows that although representatives of local governments in smart cities think that selected stakeholder or open participation are the best governance models to be implemented in their cities, they do not assign relevance in creating transparency websites or open data portals to disclose information. This result is especially true for the particular case of the offering of transparency websites. This way, the Pearson Chi-square test indicate that for both the creation of transparency websites or the creation of open data portals, RQ2 is not supported because statistical results are not significant.

By contrast, contingency tables seem to indicate an association between the offering of contact with the government and the intention to implement selected stakeholder or open participation governance models. Nonetheless, the results of the Pearson Chi-square are not significant again. Nonetheless, we can assume that the more participative model of governance, the higher offer of smart cities to contact with the city government.

Finally, similar to the result obtained for the information transparency case, there is no preference in creating e-participation platforms in the smart cities whose representatives have responded that selected stakeholder or open participation models are those to be implemented in their cities. Therefore, Pearson Chi-square does not support results and null hypothesis (RQ3) cannot be rejected.

5 Conclusions

Under the smart cities framework, the public value creation is linked to the satisfaction of citizenry's needs through the creation and use of smart technologies to increase their quality of life [1]. This research explores the relationship between the perception of smart practitioners about the model of citizen participation and the use of new technologies by city governments with the aim at improving e-participation of the citizenry in cooperative environments of smart cities. Findings based on smart city practitioners suggest the need to allowing citizens to actively participate in the city, mainly in the development of innovative ways to empower citizens and encourage civic engagement.

So, cooperation and collaboration with service stakeholders must be encouraged, within and outside organizational boundaries [20].

The behaviour expressed in this research by smart practitioners could be explained because the organizing capacity of a city government is often viewed in terms of internal organization but the creation of public value in smart cities demand local governments to advance to a higher transformative level of governance, called as smart urban collaboration [15], which is based not only on the transformation of the internal organization but also of the external organization [15].

Therefore, a main challenge in smart cities is concerned with understanding how to design tools to facilitate online deliberation and support collaborative working environments. Nonetheless, according to our results, sample smart cities in our study are not promoting active participation of citizens. In fact, our results indicate that sample smart cities do not generate virtual environments for information transparency or for favouring fluid interaction between local governments and citizens. Hypotheses testing and contingency tables confirm that smart practitioners are prone to introduce open governance models but their smart cities are not putting available technological tools for achieving that aim. Thus, it is time to take steps from "intentions" to "actions". Perhaps a more active citizenship could help to demand smart cities to be more open o their queries and participation.

To achieve this aim, it is necessary to create a culture of citizen participation in public decisions. Citizens must be also smart and problems of digital divide should be solved in all cities. Recent research has found that governments are able to advocate and educate their citizens by communicating electronically with them and thus change their behaviours and attitudes toward the society [7]. So, smart cities must not only invest on ICT infrastructure, but also on staff training, human capital/education, social and relational capital, and environmental interest as important drivers of urban growth [2]. So, future research should tackle all these problems and explore the causes why citizens are not participating in cities where all aspects of open government are accomplished. Perhaps the reason is the lack of interest in participating in decision-making processes [23], the lack of technological tools to do it or the complexity of e-participation platforms to make them involved in these processes. Also, future research could be addressed to analyse whether smart cities are promoting, or not, e-participation in a higher level than those cities that cannot be labelled as smart. This way, we could advance better in the definition of smart cities and in the axes in which they are built on.

In conclusion, findings confirm that there is a dominant belief that transformation of governance is desirable and needed to make cities smart and to create public value, which promotes the use of new technologies to adopt a more participative model of governance [15]. Nonetheless, for truly effective local governance, it is essential that public managers and politicians not only govern efficiently, but that they put available technological tools for engaging citizens in open and participative information sharing and decision-making. This is a main challenge for governments in the nearer future.

Acknowledgments. This research was carried out with financial support from the Regional Government of Andalusia (Spain), Department of Innovation, Science and Enterprise (Research project number P11-SEJ-7700).

Appendix

See Tables 1, 2, 3, 4 and 5.

Table 1. Descriptive statistics of questionnaire results.

	Descriptive statistics of the responses to the e-survey											
RQ1. Do governments promote the collaborative or participative models of governance in smart cities as a way of public value creation?	Frequency						Mean	Median	Mode	Standard deviation	Maximum	Minimum
Q1.1. The model of participation in developing the smart city in your municipality is …	No	%	Yes	%	**Total**							
(a) Exclusively the municipality. No participation of other actors	62	96.88%	2	3.13%	**64**	**100.00%**	0.0313	0	0	0.1754	1	0
(b) Selected Stakeholder participation	33	51.56%	31	48.44%	**64**	**100.00%**	0.4844	0	0	0.5037	1	0
(c) Open participation	33	51.56%	31	48.44%	**64**	**100.00%**	0.4844	0	0	0.5037	1	0

Source: Own elaboration.

Table 2. Analysis performed in this research. Variables and method of evaluation.

Area of analysis	Stages analysed	Items	Description	Type of analysis	Score
E-participation	Information transparency	Web of transparency	Existence of a specific website for information transparency	Quantitative	0/1
		Open data website	Existence of a specific Open data website for information transparency. Also, we analyze the format of the information uploaded on the website (pdf, xls, csv)	Quantitative	0/1
	Active participation	Citizen participation platform	Existence of a specific citizen participation platform. Also, we analyze whether the platform is used for public consultations, discussions and/or petitions	Quantitative	0/1
		Contact with the government	Existence of a specific space for contacting with the government of the smart city	Quantitative	0/1

Source: Own elaboration.

Table 3. Transparency websites and open data websites in total sample population and respondents to the e-survey.

	Total sample population		Respondents to e-survey		Number of information formats used	Number of SC			
	Number of SC over 85 city members	%	Number of SC over 32 respondents	%		Total population	%	Respondents to e-survey	%
Transparency websites	11	12.94%	8	25.00%	5 information formats	2	2.35%	1	3.13%
Open data websites	49	57.65%	18	56.25%	4 information formats	8	9.41%	4	12.50%
Information format					3 information formats	13	15.29%	3	9.38%
pdf	15	17.65%	4	12.50%	2 information formats	15	17.65%	6	18.75%
xls	19	22.35%	6	18.75%	1 information format	12	14.12%	6	18.75%
csv	38	44.71%	14	43.75%					
json	27	31.76%	11	34.38%					
xml	24	28.24%	8	25.00%					
Total cities with file formats in open data websites	**9**		**16**						

Source: Own elaboration.

Table 4. E-participation tools used in total population and respondents to the e-survey.

	Total population		Respondents to e-survey		Aspects dealt with into the citizen participation platform website			Number of smart cities					
	Number of SC over 85 city members	%	Number of SC over 32 respondents	%	Public consultations	Discussions	Petitions	Total population	% Total	% over 85 SC	Respondents to e-survey	% Total	% over 32 SC
Citizen participation platform website	38	44.71%	6	18.75%	X	X	X	13	**33.33%**	**15.29%**	3	**27.27%**	**9.38%**
Contact with the local government	81	95.29%	15	46.88%	X	X		4	**10.26%**	**4.71%**	1	**9.09%**	**3.13%**
						X	X	4	**10.26%**	**4.71%**	3	**27.27%**	**9.38%**
Aspects into participation platforms					X		X	2	**5.13%**	**2.35%**	0	**0.00%**	**0.00%**
Public consultation	6	7.06%			X			9	**23.08%**	**10.59%**	2	**18.18%**	**6.25%**
Discussions	6	7.06%				X		3	**7.69%**	**3.53%**	2	**18.18%**	**6.25%**
Petitions	13	15.29%					X	4	**10.26%**	**4.71%**	0	**0.00%**	**0.00%**

Source: Own elaboration.

Table 5. Contingency tables and hypothesis testing.

Attribute/E-survey model of governance			The model of participation in developing the smart city in your municipality is …				Pearson Chi-Square		
			Exclusively the municipality. No participation of other actors	Selected Stakeholder participation	Open participation	Total	Value	Df	p (2-sided)
Information transparency	The Smart City has a government transparency website	No	1	12	11	**24**	0.605[a]	2	0.738953
		Yes	0	5	3	**8**			
	The Smart City has an open government data website	No	1	8	8	**17**	1.224[a]	2	0.542178
		Yes	0	9	6	**15**			
Active participation	The Smart City offers the possibility to contact with the government	No	0	1	0	**1**	0.911[a]	2	0.634189
		Yes	1	16	14	**31**			
	The Smart City has a citizen participation platform for promoting citizen involvement in public decisions	No	1	8	10	**19**	2.597[a]	2	0.273004
		Yes	0	9	4	**13**			
		Total	**4**	**68**	**56**	**128**			

[a]2 boxes (33.3%) have an expected frequency less than 5. Minimum expected frequency is 0.41
Source: Own elaboration.

References

1. Berntzen, L., Johannessen, M.R.: The role of citizen participation in municipal smart city projects: lessons learned from Norway. In: Gil-Garcia, J.R., Pardo, T.A., Nam, T. (eds.) Smarter as the New Urban Agenda. PAIT, vol. 11, pp. 299–314. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-17620-8_16
2. Caragliu, A., Del Bo, C., Nijkamp, P.: Smart cities in Europe. J. Urban Technol. **18**(2), 65–82 (2011)
3. Considine, M., Lewis, J.: Governance at ground level: the front-line bureaucrat in the age of markets and networks. Public Adm. Rev. **59**(6), 467–480 (1999)
4. Dameri, R.P., Negre, E., Rosenthal-Sabroux, C.: Triple Helix in Smart cities: a literature review about the vision of public bodies, universities, and private companies. In: 2016 49th Hawaii International Conference on Systems Sciences (HICSS), Hawaii, USA, pp. 2974–2982. IEEE (2016)
5. EUROCITIES: Eurocities Strategic Framework 2014–2020. Towards an urban agenda for the EU. Eurocities, Brussels (2016)

6. EUROCITIES: About us (2017). http://www.eurocities.eu/eurocities/about_us. Accessed 1 Feb 2017
7. Feeney, M.K., Brown, A.: Are small cities online? Content, ranking, and variation of US municipal websites. Gov. Inf. Q. **34**(1), 62–74 (2017)
8. Florida, R.: The Rise of the Creative Class: And How it is Transforming Work, Leisure, Community and Everyday Life. Basic Books, New York (2002)
9. Gascó, M.: Living labs: implementing open innovation in the public sector. Gov. Inf. Q. **34** (1), 90–98 (2017)
10. Giffinger, R., Gudrun, H.: Smart cities ranking: an effective instrument for the positioning of the cities? ACE: Archit. City Environ. **4**(12), 7–26 (2010)
11. Gil-Garcia, J.R., Pardo, T.A., Nam, T.: A comprehensive view of the 21st century city: smartness as technologies and innovation in urban contexts. In: Gil-Garcia, J.R., Pardo, T.A., Nam, T. (eds.) Smarter as the New Urban Agenda. PAIT, vol. 11, pp. 1–19. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-17620-8_1
12. Khodabakhsh, P., Fathi, H., Mashayekhi, S.: Planning for future urban services in the smart city era: integrating e-services in urban planning process. Armanshahr Archit. Urban Dev. **9** (16), 153–168 (2016)
13. Lee, S.: Creative citizenship and the public policy process: a flibbertijibbet, a will-o-the-wisp, a clown? Cult. Sci. J. **8**(1), 85–90 (2015)
14. Mainsah, H.N.: Social media, design and creative citizenship: an introduction. Digit. Creat. **28**(1), 1–7 (2017)
15. Meijer, A., Rodríguez Bolívar, M.P.: Governing the smart city: a review of the literature on smart urban governance. Int. Rev. Adm. Sci. **82**(2), 392–408 (2016)
16. Moore, M.H.: Creating Public Value: Strategic Management in Government. Harvard University Press, Cambridge (1995)
17. Pierre, J.: Comparative urban governance: uncovering complex causalities. Urban Aff. Rev. **40**(4), 446–462 (2005)
18. Rodríguez Bolívar, M.P.: The influence of political factors in policymakers' perceptions on the implementation of Web 2.0 technologies for citizen participation and knowledge sharing in public sector delivery. Inf. Polity **20**(2), 199–220 (2015)
19. Rodríguez Bolívar, M.P.: Characterizing the role of governments in smart cities: a literature review. In: Gil-Garcia, J.R., Pardo, T.A., Nam, T. (eds.) Smarter as the New Urban Agenda. PAIT, vol. 11, pp. 49–71. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-17620-8_3
20. Rodríguez Bolívar, M.P.: Policy makers' perceptions on the transformational effect of Web 2.0 technologies on public services delivery. Electron. Commer. Res. **17**(2), 1–28 (2017)
21. Rodríguez Bolívar, M.P., Alcaide Muñoz, L., López Hernández, A.M.: Scientometric study of the progress and development of e-government research during the period 2000–2012. Inf. Technol. Dev. **22**(1), 36–74 (2016)
22. Gramberger, M.: Citizens as Partners OECD Handbook on Information, Consultation and Public Participation in Policy-Making. Organization for Economic Cooperation and Development, Paris (2001)
23. Hibbing, J.R., Theiss-Morse, E.: Stealth Democracy: Americans' Beliefs About How Government Should Work. Cambridge University Press, Cambridge (2002)
24. Rodríguez Bolívar, M.P.: Governance models for the delivery of public services through the web 2.0 technologies: a political view in large Spanish municipalities. Soc. Sci. Comput. Rev. **35**(2), 203–225 (2017)
25. European Commission: Europe for citizens. Brussels, version valid as of 2017 (2017)

26. Denters, B.: Participation and democratic accountability: making a difference for the citizens. In: The Future of Local Government in Europem, pp. 79–100. Nomos Verlagsgesellschaft mbH & Co. KG (2017)
27. Snedecor, G.W., Cochran, W.G.: Statistical Methods. The Iowa State University Press, Iowa (1981)
28. Sprent, P., Smeeton, N.C.: Applied Nonparametric Statistical Methods. Chapman & Hall/CRC Taylor & Francis Group LLC, Boca Raton (2007)

Regulatory Compliance and Over-Compliant Information Sharing – Changes in the B2G Landscape

Bram Klievink(✉), Marijn Janssen, Haiko van der Voort, and Sélinde van Engelenburg

Delft University of Technology, Delft, The Netherlands
{A. J. Klievink, M. F. W. H. A. Janssen, H. G. vanderVoort, S. H. vanEngelenburg}@tudelft.nl

Abstract. Business-to-government information exchange has over the past decades greatly benefited from data exchange standards and inter-organisational systems. The data era enables a new shift in the type of information sharing; from formal reporting to opening up full (and big) data sets. This enables new analytics and insights by government, more effective and efficient compliance assessment, and other uses. The emphasis here shifts from establishing formats to deciding what information can be shared, under what conditions, and how to create added value. There are numerous initiatives that explore how to put data to better use for businesses, for government and for their interactions. However, there is limited attention to exactly how these new forms of extensive data sharing affects the supervision relationships. In this paper, we exploratively look across three research projects to identify the implications of information sharing beyond the regulatory requirements ('over-compliant'). We find that the lack of attention to those implications lead to solutions that are hard to scale up and present unexpected consequences down the line, which may negatively impact the future willingness to explore new potential added value of data sharing.

Keywords: Business-to-government · B2G · Regulatory compliance Supervision · Information sharing

1 Introduction

Ever since their inception, information and communication technologies (ICTs) have played a big role in inter-organisational information exchange. We are here concerned with a specific domain for information exchange; that between businesses and government (or B2G information exchange). Electronic B2G information exchange is useful for all kinds of data that businesses must provide to government, for example for supervision purposes, taxation and statistics. Companies must comply with many laws and regulations and the information they collect and hold can be vital in demonstrating compliance to government actors and other stakeholders [1, 2]. ICTs help companies collect evidence about compliance and support the exchange and evaluation of that evidence. The added efficiency that this brings, has been one of the main drivers for interoperability, standards and inter-organisational systems in this domain. To

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 249–260, 2018.
https://doi.org/10.1007/978-3-319-98690-6_21

demonstrate compliance, the way that companies themselves collect and process information requires certain controls and standardised formats and access by government [3, 4]. Such B2G reporting is highly regulated, with obligations pertaining to scope, scale, timing and format for sharing. Standardised data formats and interfaces as well as regulatory and supervisory instruments such as EDP auditing, play an important role in how ICTs assist companies in demonstrating compliance - and governments in assessing it. The introduction of electronic data exchange several decades ago, can be seen as the tipping point for digital B2G exchange. Many e-government systems that are meant to collect or receive business information accept documents according to standardised data and messaging formats. These highly structured, yet age-old formats are still widely used due to their large installed base. Yet, we can now witness a new tipping point based on the increasing amount of data and the technical capability to share and act on original source data in near real time.

There is a host of enterprise information systems (EIS), monitoring systems and inter-organisational systems (IOS) that all collect, produce and/or store information. This information is not just key to the operations of the company, governments can also use the information. Scholars, companies and government are now working on new ways of information sharing, that are based on applying the same formats and systems for both business operations and B2G, instead of B2G exchange being a derivative of the full source data as they exist in the enterprise systems [1, 2]. This allows piggy-backing on original business data by governments [5] and that support new value creation by going beyond the formal reporting and compliance relationship [6]. The emphasis then shifts from the old practice of identifying formal data elements and reports, to identifying and deciding which information may or should be shared to create value for regulatory supervision and/or other purposes. It may even enable governments to do data analytics and identify patterns that they otherwise could not. We summarise this as "over-compliant" information sharing, by which we mean that businesses open up more information than that they are legally required to.

Although over-compliant B2G information sharing is technically feasible and looks promising, a significant part of the 'business case' is related to innovation in supervision that are enabled by it. A prominent example here are companies that share original business records with more information than is formally required, who may be 'rewarded' with less inspections or a lower administrative burden. However, the work on such innovations do mention but rarely focus on the implications of additional data sharing on the compliance and supervision relationship between the parties involved. The work that does address this topic, often does that from the perspective of a single case or project in which over-compliance (our term) plays a role. This paper aims to address this gap by exploring three projects on over-compliant B2G information sharing and what that means for the relationship for the relationship between the companies and supervision or oversight bodies. To this end, we study three research projects on over-compliant information sharing and look for insights across those cases to identify the implications that such forms of sharing may have on B2G information exchange. Specifically, we do that from a perspective of regulatory compliance based on B2G information sharing.

The paper is structured as follows: in the next section, we discuss the current paradigm of B2G information sharing, which heavily relies on inter-organisational

systems and standardised information exchange. We also there discuss the recent literature on over-compliant information sharing. In the third section, we briefly explain our approach and the three projects we study. In Sect. 4, we present our findings. Finally, in Sect. 5, we summarise those in terms of conclusions.

2 Background

There are many considerations for businesses to (not) share data with government. For instance, technical considerations and capacity considerations, the risks of data ending up with competitors, and the risk of sharing inaccurate data seem to be relevant. From a more political viewpoint, being transparent towards a regulator might put the business in a vulnerable position for the obvious reason that regulators may have enforcement instruments that may hurt the business. This way of opening up may involve risks to the business and the way businesses want to manage these risks are relevant for regulators seeking compliance. As we are interested in the relationship between over-compliant B2G information sharing and regulatory governance, in this section we discuss the relevant background of both. We first dive into how IT supports and changes this relationship. The main part of this section follows after that: we discuss regulatory governance literature to find the main context variables that affect compliance and the relationship between supervision agencies and businesses.

2.1 Information Sharing

The introduction of information systems has greatly transformed the practice of information exchange between businesses and governments. Governments have put in place information systems for exchanging data with companies, including data for taxation, business reporting, statistical purposes, and for establishing compliance to relevant policies and regulation. Many of these systems are directly relevant for operational processes at both the business and the government side. Hence, their operations become mission-critical to at least some of the organisations involved. Pressing impediments and project risks are salient and need to be addressed [7–9]. The antecedents of use of information systems has been a topic of interest for a few decades now [10]. Among the key factors that influence adoption and diffusion of IOS, are the characteristics of the system, their benefits, institutional forces, resource dependence and readiness [10]. The implications on the organisation are often profound and may require restructuring of the organisation and changes in and standardisation of the information function in the organisation to adapt to inter-organisational information sharing landscape. As ICT may reinforce dependencies, organisations may be hesitant to deeply integrate them in their own systems, for example to avoid a lock-in by partners [11]. Trust in sharing partners is key to reap full benefits yet is difficult to develop. Even though this may be less of a problem in B2G information sharing, at least some of the information shared with governments will depend on information of (trading) partners and also the risks and benefits may be at the network level instead of at the level of the individual organisation [12]. In sum, B2G information sharing used

to be mainly about standardised information exchange via interoperable systems, to enhance efficiency in reporting and processing data.

In the current day and age, the information flow from business to government can be based directly in business information systems or allowing governments to tap into information flows that are used for multiple purposes [1, 5]. To enable this shift, internal and external control over companies' operations are crucial. As many operations are at least supported by information systems, the control objectives also concern the information systems [3]. Forms of over-compliant B2G information sharing, as presented in the introduction [5, 6] present challenges that may alter the relationship between regulator or supervision agency and the company providing the information. Studies that touch upon over-compliant information sharing often acknowledge that this is the case, yet rarely cover this change thoroughly. The benefits are often phrased as: re-using data collected for other purposes saves costs, sharing more information allows supervision bodies to make better decisions which would lead them to not burden over-compliant companies, and there might also be reputational gains [3, 13, 14]. At the same time, new incentive schemes are surely required, as companies typically incur costs and are susceptible for additional risks if they give more detail than necessary. But then what does over-compliant information sharing mean for compliance and the supervision by governments?

2.2 Compliance and Supervision

This subsection covers the literature on regulatory compliance and takes a political view. In the literature, we distinguish five context variables: (1) Compliance motives of the regulated, (2) Compliance enforcement strategies of the regulator, (3) Compliance relationships, (4) Public compliance regimes, (5) The broader institutional context.

Compliance motives of the regulator is a classic concern in regulatory literature. Kagan and Scholz made an authoritative typology of regulated businesses informed by both their motives and capabilities to comply [15]. The typology includes "amoral calculators" and "political citizens" as businesses that under some conditions are non-compliant. The first is informed by their own interest and the second with the moral ambivalence or regulation [16]. A third type of non-compliant business as identified by Kagan and Scholz are the "organisational incompetent", being those that might want to comply, but are not able to. The willingness and ability to comply may not only determine the eagerness to open up. The motives also suggest that the decisions what to share and how to share it are in their core of a strategic kind. If these decisions are perceived this way, data and sharing methods become strategic assets.

Compliance enforcement strategies. From the side of the regulator the enforcement style is relevant to the tendency of businesses to share data. In regulatory governance literature two main styles are distinguished [16–18]. A first is the classic style of enforcing the law if a business violates it. This style is usually coined as deterrence [15, 19], penalism [20], coercion or the sanctioning style [21]. After incidents a call for stricter enforcement is commonly heard, however this strict style is plagued by principal- agent problems since regulators have to rely on data to detect law violations [18, 21]. An alternative style is a coaching style that is more focused on learning by the business, better relations between regulator and business and development of

professional norms to cope with moral ambivalence of rules [18]. These regulatory styles influence the strategic considerations for businesses to share data. The main consideration here is the probable impact of data about law violations on the business. If data show law violations and these violations will automatically result in public interventions such as fines, the regulated wouldn't be too eager to open up.

As a function of both the motive of the business and the style of the regulator, their compliance relationship is a context factor of considerable importance. De Bruijn et al. [21] found a seemingly endless amount of strategies both the regulator and the business have to confuse the other. Since the nineties, regulatory governance literature started to focus on the questions how the relationships between regulator and business can be made more productive – at least for the regulator. Typical topics are the responsiveness of regulators to the behaviour and motives of the businesses [22], the way regulators interact with their regulatory counterparts within business such as compliance managers (Parker) and the way trust between regulators and businesses evolves [23, 24]. The relevance of these more relational issues is their capacity to make the behaviours around data exchange some more predictable. Some more stability of expectance about what is being done with the data may encourage opening up in the course of time.

A fourth variable is about regulatory policies rather than regulatory styles. On this level, public compliance regimes define the inspector's job. What are the rules the businesses have to comply with? May distinguishes prescriptive regulation, performance-based regulation and system-based regulation [25]. They mainly refer to what the rules prescribe, which are actions, results and processes. These regimes matter for data exchange for at least two reasons. First, per regime, different types of data are needed and some data are more sensitive than others. Second, they imply different responsibilities of the business. For instance, performance-based regimes leave leeway to how results or outcomes are reached [26]. System-based regimes usually refer to management systems the businesses control themselves, and this way aim to reinforce the self-regulatory capacity of the regulated business. As such, these regimes – more than others - imply an own responsibility for businesses to collect compliance data and also imply a role for regulators to drive away from this data collection process and collect data on a metalevel (i.e. is there a management system in place?).

Finally, the broader institutional context of the relation between regulator and industry matter for decisions to share data. The relation between regulator and business doesn't develop in isolation. Regulated industries face many different public and private actors demanding responses, including banks, NGO's, insurance companies, trade organisations, and governments [21, 27]. In the 2000s a 'decentred view' on regulation became in vogue. With such a view government is no longer perceived to act as the central regulator of the public sphere. Regulation is essentially not state-centred, but rather a result of various public and private regulators seeking to impose rules to others [28–30]. From this viewpoint, Black and Baldwin introduced the idea of 'really responsive regulation' [31]. Regulators would not only respond to compliance of regulated firms, but also to their institutional environments, interactions of regulatory controls and change – among other aspects [31, 32]. Data exchange cannot be isolated from its context. If perceived so nonetheless, special effects may happen. For example, media might be eager to publish about business performances based on data offered to governments, sometimes devoid from any nuance [33].

3 Approach and Project Descriptions

This paper is concerned with the implications of over-compliant B2G information sharing on compliance and supervision. Recent work that does address this topic, often does that from the perspective of a single case or project in which over-compliance (our term) plays a role. In this paper, we seek to go beyond that and look across three of such projects to extract a more comprehensive view on the issues that over-compliant B2G information sharing presents for the compliance relationship between government and companies.

The three projects that we focus on are research or research & development projects. That makes sense, as implementing these solutions in practice lead to numerous challenges and issues. The three projects are: SBR (a public-private initiative in the Netherlands), CASSANDRA (an FP7 project), and JUST (a Dutch research project). The first two were active in the beginning of this decade, the latter is an ongoing project. All of them concern the Dutch situation, with the exception of Cassandra, which operated in various European countries (the Netherlands included). Given the explorative nature of this study, we did not employ a rigorous comparative design, yet revisit the documents and papers on these projects to extract what can be learned from them about the supervision relationship. The authors were involved in these projects [1, 2, 6, 12, 34, 35], which hurts replicability but we do trade that off with much in-depth, first-hand information. The study is thus based on participatory observation by the authors in these three projects as well as on the documentation in formal project deliverables and papers published on the projects.

3.1 Description of the Projects

SBR: Standard Business Reporting. The first initiative we reflect on is a Dutch initiative called Standard Business Reporting (SBR). It was developed and applied in the Netherlands to change B2G reporting [1, 2]. The specific issue addressed by SBR is that often companies have to provide similar (although not always the same) information to multiple government agencies, each with their own systems, formats and definitions. This leads to multiple interfaces, and checks on data elements, standards and definitions. Consequently, information is sometimes shared through separate reports (IOS, e-mail), making the extraction of key information a time-consuming and error-prone process. The SBR project delivered a platform that builds on the eXtensible Business Reporting Language (XBRL). A key contribution of the project was to agree on standardization of data (syntax and semantics) [2, 6].

CASSANDRA. Similar to SBR, the many actors that are involved in international trade also report much information to various government agencies as well as to other parties in the chain. As international trade is typically organised in supply chains involving many companies, there are many handovers between those actors, even before the information is supplied to the government by one of them. The information is therefore often fragmented and information quality can be poor [34, 36]. This project focused on leveraging IT innovations to improve the information exchange between actors worldwide by creating electronic connections between organisations. In the project, the systems of supply chain partners are interconnected and jointly formed an

international information infrastructure [5, 34, 37]. Through the infrastructure, data can be shared among supply chain partners as well as with government. Government agencies involved in the supervision of international trade (e.g. customs, food and product safety, tax) can get a better view on the goods actually being traded and entering the country if they can get detailed information from the original source information systems at all of the parties involved in the shipment [6].

JUST. The JUST (JUridical and context-aware Sharing of informaTion for ensuring compliance) project is related to the infrastructures of the aforementioned projects yet focuses on an important development: context-awareness. The other projects did cover important issues on governance, collaboration and trust. However, they typically assume it is a onetime decision for companies; they either stick to formally reporting the information they are required to share, or they decide to open up more data than they are legally required to do. However, the developments towards context-aware systems mean that information gets shared. What that means, may depend on the context. That gives rise to a specific challenge with potentially great implications for B2G information sharing: context information about the requester of access to data, together with relevant business rules, can play a key role in the decision about whether to share data [35]. Also the legality of the sharing of information can depend on various context variables, including the circumstances, jurisdiction, applicable regulations, original source of the data, business relationship between the company and the source, public interest considerations, and many other factors.

Businesses have to take these and other matters into account when deciding to share data with governments beyond their formal obligation. That means that sharing is not a decision that is made for all data and all time, but there are many factors that play a role and those factors may play out differently in different instances. For example, the information that a company shares from the operations in one supply chain may be in much greater detail, than what the same companies shares on operations in another supply chain (with other partners and data from different sources).

4 Findings: Compliance Challenges and Consequences for Supervision

The three projects each paint different pictures of the effects of over-compliant B2G information sharing on supervision. In brief, in SBR big challenges can be found in the interaction among government agencies, in Cassandra we find risks of getting punished for ‘good behaviour’ and in JUST we find that the technical developments towards context-awareness may have some implications. In this section, we discuss these findings, albeit a bit briefly.

Intra-governmental issues. The concept is often presented as an innovation in public-private interactions. In the SBR case, the core components were quickly considered public infrastructure and hence the responsibility of government. That would provide continuity and stability that is needed to make the change systemic to the way businesses store and report information. Furthermore, for some of the services based on the innovation, the law just states that they are the government’s responsibility. Yet,

several market representatives were involved in standard selection, taxonomy creation and decision-making. Apart from this being a challenge on the public-private interface, also within government there are many agencies and other organisations that act as supervisor on a specific area. Many companies have dealings with multiple of them (e.g. tax, customs, food and product safety, financial regulators, etc.) and the file-once principles requires that they are all able to act on the same data, using the same standards, following the same procedures. The lead agency in the project (the tax administration) thus had to make many decisions that would work for them, but also for the companies and other agencies. An important challenge here is what if others decide not to follow, for example because their task is less information intensive? The differences between government agencies is not trivial; legislation does not allow to re-use data collected for one purpose to be used for different purposes. Furthermore, as various agencies have different legal bases (e.g. tax or commercial) for their data requests, reports may use the same data but end up with different interpretations. These issues were also found in the Cassandra case.

Interoperability and openness of data is obviously a key issue when it comes to the information system aspects of the innovation. Yet, there are also major implications for compliance. In the cases we find situations in which data originally comes from third parties (i.e. the reporting company receives this data from a third party as part of their operations). In our cases that concern international trade data, data often gets updated in the business systems, for example as quantities or destinations change. In over-compliant information sharing, supervisors have access to the company data. However, in regulations, there are formal reporting moments (for example entry summary declarations for incoming goods); if data get updated after that formal reporting moment, what is then the status of that information? The update itself is information for the supervisor (both the update and the fact that is was updated), but does the regulatory regime allow the agency to act on that? And what does that mean for decisions based on the earlier data?

There are ***different speeds***, which is especially visible in the SBR case. The first phases were relatively experimental and first movers faced high transaction costs. There were many incentives for businesses to wait and see, as old alternatives may be not efficient on paper, yet were in practice. It is attractive for them to free ride on the investments of first movers. It was to government to show the efficiency of SBR and make it attractive to step in as quick as possible. A catch 22 situation proved to be a risk: the project needed a critical mass of business to participate to mature, while businesses waited SBR to mature. Also among regulators some there were front runners and others were laggards. As a result, front runners are exposed to risk of failure. At the same time, the standards and customs developed by front runners may become de facto standards. The Dutch Tax Administration, as the main proponent of SBR, didn't always wait for other government agencies to participate. For instance, they developed their own system-to-system channel (BAPI), which reduced possibilities to find a collectively rational solution.

The disadvantages. Companies tend to focus on the benefits they might get when sharing additional data. Also for the government agencies, providing supervision-related benefits is key to incentivise companies to open up their information systems. In

practice, however, we found that it is not uncommon that the data that is shared beyond their duty to share, at times also makes errors visible, such as underreporting or misreporting. These errors might not have been detectable under the 'normal' supervision regime, but now lead to fines precisely because of the over-compliance of the company. In the instances where we found this, the companies did not want to be punished for something that is a result of their attempt to be better than others. Also the supervision agencies would rather reward the companies than punish them. Yet, if an inspector finds something wrong, they have no choice but to act on that information, even if provided voluntarily. This shows that the compliance motives of the company are supported by the relationship with the supervisor and the way that over-compliance is reflected in the enforcement strategy, yet this spirit of collaboration at times finds itself at odds with the formal regime.

To follow up on the ***regulatory regime and style***, the regulatory regime will have to allow for systems-based supervision to enable the benefits. Whether such supervision regime is possible depends on the data (transaction data, output data, meta data on the organisation) that is available and the extent to which supervisors depend on that kind of data of the regulated. Furthermore, as over-compliance is often new and sometimes restricted to the 'best in class', the style of the supervisor is often based on collaboration. This means that the supervisor will coach the company towards a situation that they are both happy with. This is also an intensive phase; although actual supervision for that company will over time take less resources, setting everything up, takes up more. The question is for what number of companies that can be done and what that would require of auditors. It also depends on the specific sector. For example, are there professional incentives for self-regulation?

Finally, ***context-awareness*** is an issue, especially for sensitive data. The new possibilities to process large amounts of data (e.g. data mining) make it harder to protect sensitive data. A data element is not necessarily sensitive in and by itself, but may become that when aggregated or combined with other data elements. It is challenging to keep track of which combinations are sensitive and who has (had) access to what data in case there are many parties involved, or there are large volumes of data with many different elements. To still protect sensitive data, the juridical and technical safeguards need to be aligned. That depends on the motives of both the regulated and the supervisor, but also on the institutional context and supervision regime.

Being aware of the context in which information is created and/or shared, is important for assessing the compliance of those companies. With the abundance of information available today, there are more data and meta-data available to feed (latent) variables that measure compliance. However, this situation also leads to new challenges and questions; for example on regulatory options for taking into account context as meta-data. Furthermore, if information sharing systems themselves become more context aware, how to know what is not being shared, what is altered and what is only valuable for a specific context? What is the role and value of internal control in those situations, and what does this mean for IT auditors? If information is not shared; how can government agencies know whether that is for good contextual reasons or when that is because of unwillingness to share? How can you know what information is being withheld, and assess whether that means something for establishing compliance or providing the benefits associated with over-compliant information sharing?

Although recently studies have started to cover the technical challenges of such context-aware information sharing [38], what that means in terms of the supervision relationship, has not yet been explored. This is an important research direction, as for the further development of context-aware, over-compliant information sharing, it is vital that the supervision implications of sharing decisions are known.

5 Conclusions

In this paper we have looked at three R&D projects to extract lessons on the implications that over-compliant B2G information sharing may have on the relationship between regulatory supervision and the regulated. In the cases we find that there are many variables and complexities that play a role in that relationship which are affected by the technical innovation. The work on the technical innovations tends not to focus too much on those. Especially regulatory benefits are easily counted on as a key incentive for companies to change the relationship with the supervisor into a much more collaborative one, in which they can share additional data in return for compliance benefits. However, as this is a field where the technical innovation meets soft variables, the effects of the innovation on the relationship and vice versa may not be so clear-cut. Apart from the known issues in public-private collaboration (e.g. control, autonomy, amoral behaviour, conflicting goals and interests), even collaborations that are genuinely committed to making over-compliant B2G information sharing work for both sides encounter push-back from the economic rationality for the company and the institutional and regulatory environment of the supervision regime. A changing relationship between regulator, supervisor, and the regulated, especially when enabled by data, will have to start from certain anticipation effects, such as trust and expectations. Yet, the challenges that we found in this paper (although probably only a small portion of all those out there) show that the institutionalisation of the previous relationship when combined with uncertainties of how new technologies will play out, make the process much more cumbersome than most parties set out with.

The three projects covered in this paper have led to new supervision relationships, but only for a handful of companies and also government agencies are finding it hard to scale up. That is not strange, as even in face of great benefits, the new challenges are equally great. We have institutionalised so much value in the current supervision regimes, that we are only beginning to find out how to address things like internal control, external audits, control frameworks, discretionary freedoms, strategic selection, and many others have to be adapted to ever more data.

References

1. Bharosa, N., et al.: Tapping into existing information flows: the transformation to compliance by design in business-to-government information exchange. Gov. Inf. Q. **30**, 9–18 (2013)
2. Bharosa, N., van Wijk, R., de Winne, N., Janssen, M.: Challenging the Chain: Governing the Automated Exchange and Processing of Business Information. IOS Press, Amsterdam (2015)

3. van der Pligt-Benito Ruano, S., Hulstijn, J.: Governance and collaboration in regulatory supervision: a case in the customs domain. Int. J. Electron. Gov. Res. **13**, 34–52 (2017)
4. Veenstra, A.W., Hulstijn, J., Christiaanse, R., Tan, Y.-H.: Information exchange in global logistics chains: an application for model-based auditing (2013)
5. Tan, Y.-H., Bjørn-Andersen, N., Klein, S., Rukanova, B.: Accelerating Global Supply Chains with IT-Innovation. ITAIDE Tools and Methods. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-15669-4
6. Klievink, B., Bharosa, N., Tan, Y.-H.: The collaborative realization of public values and business goals: governance and infrastructure of public–private information platforms. Gov. Inf. Q. **33**, 67–79 (2016)
7. Van Veenstra, A.F.: IT-Induced Public Sector Transformation (2012)
8. Urciuoli, L., Hintsa, J., Ahokas, J.: Drivers and barriers affecting usage of e-customs — a global survey with customs administrations using multivariate analysis techniques. Gov. Inf. Q. **30**, 473–485 (2013)
9. Gil-garcia, J.R., Chengalur-Smith, I., Duchessi, P.: Collaborative e-government: impediments and benefits of information-sharing projects in the public sector. Eur. J. Inf. Syst. **16**, 121–133 (2007)
10. 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**, 497–518 (2008)
11. Hart, P., Saunders, C.: Power and trust: critical factors in the adoption and use of electronic data interchange. Organ. Sci. **8**, 23–42 (1997)
12. Klievink, B., Lucassen, I.: Facilitating adoption of international information infrastructures: a living labs approach. In: Wimmer, M.A., Janssen, M., Scholl, H.J. (eds.) EGOV 2013. LNCS, vol. 8074, pp. 250–261. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40358-3_21
13. Zuidwijk, R.A., Veenstra, A.W.: The value of information in container transport. Transp. Sci. **49**, 675–685 (2015)
14. Bharosa, N., Janssen, M., Hulstijn, J., van Wijk, R., de Winne, N., Tan, Y.: Towards a lean-government using new IT-architectures for compliance monitoring. In: Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance - ICEGOV 2011, p. 147 (2011)
15. Kagan, R.A., Scholz, J.: The criminology of the corporation and regulatory enforcement strategies. In: Blankenburg, E., Lenk, K. (eds.) Organisation und Recht. Jahrbuch für Rechtssoziologie und Rechtstheorie, vol. 7, pp. 352–357. Springer, Heidelberg (1980). https://doi.org/10.1007/978-3-322-83669-4_21
16. Hawkins, K.: Environment and Enforcement, Regulation and the Social Definition of Pollution. Clarendon Press, Oxford (1984)
17. Braithwaite, J.: To Punish or Persuade: Enforcement of Coal Mine Safety (1985)
18. Sparrow, M.: The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance (2011)
19. van Wingerde, K.: The limits of environmental regulation in the global economy. In: van Erp, J., Huisman, W., Vandewalle, G. (eds.) The Routledge Handbook of White-Collar and Corporate Crime in Europe, pp. 260–275. Routledge, London (2015)
20. Reiss, A.: Consequences of compliance and deterrence models of law enforcement for the exercise of police discretion. Law Contemp. Probl. **47**, 83 (1984)
21. de Bruijn, H., ten Heuvelhof, E., Koopmans, M.: Law Enforcement: The Game Between Inspectors and Inspectees (2007)
22. Ayres, I., Braithwaite, J.: Responsive Regulation: Transcending the Deregulation Debate (1992)

23. Six, F.E., Verhoest, K.: Trust in regulatory regimes; scoping the field. In: Six, F.E., Verhoest, K. (eds.) Trust in Regulatory Regimes, pp. 1–36. Edward Elgar, Cheltenham (2017)
24. Van Der Voort, H.: Trust and cooperation over the public-private divide. In: Six, F.E., Verhoest, K. (eds.) Trust in Regulatory Regimes, pp. 181–204. Edward Elgar, Cheltenham (2017)
25. May, P.: Regulatory regimes and accountability. Regul. Gov. **1**, 8 (2007)
26. Coglianese, C.: Management-based regulation: prescribing private management to achieve public goals. Law Soc. Rev. **37**, 691 (2003)
27. Gunningham, N., Grabosky, P., Sinclair, D.: Smart Regulation: Designing Environmental Policy (1998)
28. Black, J.: Decentring regulation: understanding the role of regulation and self-regulation in a "post-regulatory" world. Curr. Leg. Probl. **54**, 103 (2001)
29. Haines, F.: The Paradox of Regulation: What Regulation Can Achieve and What It Cannot (2011)
30. Garcia Martinez, M., Verbruggen, P., Fearne, A.: Risk-based approaches to food safety regulation: what role for co-regulation? J. Risk Res. **16**, 1101 (2013)
31. Baldwin, R., Black, J.: Really responsive regulation. Mod. Law Rev. **71**, 59 (2008)
32. Black, J., Baldwin, R.: Really responsive risk-based regulation. Law Policy **32**, 181 (2010)
33. Van der Voort, H., Kerpershoek, E.: Measuring measures: introducing performance measurement in the Dutch health care sector. Pub. Money Manag. **30**, 63–68 (2010)
34. Klievink, B., et al.: Enhancing visibility in international supply chains: the data pipeline concept. Int. J. Electron. Gov. Res. **8**, 14–33 (2012)
35. Van Engelenburg, S., Janssen, M., Klievink, B.: Design of a software architecture supporting business-to-government information sharing to improve public safety and security: combining business rules, events and blockchain technology. J. Intell. Inf. Syst. 1–24 (2017). https://doi.org/10.1007/s10844-017-0478-z
36. Hesketh, D.: Weaknesses in the supply chain: who packed the box? World Cust. J. **4**, 3–20 (2010)
37. Baida, Z., Rukanova, B., Liu, J., Tan, Y.: Preserving control in trade procedure redesign – the beer living lab. Electron. Mark. **18**, 53–64 (2008)
38. van Engelenburg, S., Janssen, M., Klievink, B.: What belongs to context? In: Cerone, A., Roveri, M. (eds.) SEFM 2017. LNCS, vol. 10729, pp. 101–116. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74781-1_8

Artificial Intelligence, Data Analytics and Automated Decision-Making

Using Geocoding and Topic Extraction to Make Sense of Comments on Social Network Pages of Local Government Agencies

Pedro C. R. Lima[1], Raissa Barcellos[2(✉)], Flavia Bernardini[2], and Jose Viterbo[2]

[1] Institute of Science and Technology – ICT, Fluminense Federal University – UFF, Rio das Ostras, RJ, Brazil
pedroccrl@gmail.com
[2] Institute of Computing – IC, Fluminense Federal University – UFF, Niteroi, RJ, Brazil
raissabarcellos@id.uff.br, {fcbernardini,viterbo}@ic.uff.br

Abstract. Social networks have become an important channel for exchanging information and communication among citizens. Text mining, crowdsourcing and data visualization are some approaches that allow the information and knowledge extraction from texts in comment formats, exchanged between citizens in social networks. This movement can be indirectly used as a bias for popular participation, gaining prominence in the construction of smart cities. The objective of this work is to present a method that geocodes citizens' comments made on posts in Social Network Pages of Local Government Agencies, and extracts the most frequent topics present in these comments. In order to validate our method, we implemented a web system that implements the steps of the proposed method, and conducted a case study. The tool, and consequently the steps of the presented method, was evaluated by four software developers, which indicated that the tool was easy to use, new knowledge could be extracted from it, and some interesting improvements were pointed out by them.

Keywords: Social networks · Data analysis · Information visualization
Smart cities · E-Participation

1 Introduction

The Smart City concept still draws much attention when it defines urban development policies and popular participation. In this context, Terán, Kaskina and Meier [1] propose a maturity model for Cognitive Cities, strongly based on a maturity model for electronic government, which values popular participation in government decision-making processes. Dameri and Rosenthal-Sabroux [2] discuss the issue of building public value by citizens, which can only be acquired by popular participation. Currently, social networks offer a range of information, provided by citizens, on pages of

P. Parycek et al. (Eds.): EGOV 2018, LNCS 11020, pp. 263–274, 2018.
https://doi.org/10.1007/978-3-319-98690-6_22

local governments. However, this source of information is still little explored to understand the quality of the provided services, and leverage new public services.

We could observe several posts by citizens on social networks, specifically in some local government agencies pages, suggesting or complaining about public services provided to citizens. Reddick, Chatfield and Ojo [3] present a conceptual framework for using social media text mining analytics and visualization e-participation. Their study case showed that using analytics techniques for mining Social Networks to gather information from citizens are quite useful. So, exploring techniques for detecting frequent mentioned topics in social network pages of local government agencies is an interesting approach. Also, data visualization has been explored in many works for leveraging open data exploration and interpretation [4, 6], which can also be used in this scenario.

In this work, we propose a method to analyze user's comments, published on a social network page of a local government agencies. We premised that some comments on posts in such pages often contain comments about locations in a city that are not necessarily related to the related post. The method comprises the identification of (i) the locations (streets, avenues and others) mentioned in the comments, to display those comments in maps; and (ii) the most frequent topics. For both tasks, we used techniques from Text Mining area for automatic geocoding the comments and automatic extraction of the topics. Geocoding comments is an interesting approach even when the comments has metadata indicating the location it was posted, because not necessarily the comment is related to the place where the user is located when posting the message. For instance, someone could comment, when arrives at home, that on the street where he/she works there is a large hole. Tools like Google Maps API are able to show points in a map given the zip code. Hence, we propose to geocode comments by linking the comment to the zip code of the location mentioned in the comment. For this task, a dataset containing the name of the streets, its respective neighborhoods and its zip code is needed. It is important to observe that this method comprises the first step for acquiring new information and knowledge from social networks related to citizens claims. In order to evaluate the proposed method, we developed a web system that shows the map visualization, with points representing comments that refer to the respective location; and the most frequent identified topics mentioned in the processed page. In addition, we also conducted an experimental analysis of the method using the page of a governmental agency from a Brazilian city.

This paper is organized as follows: Sect. 2 presents the theoretical background on Smart Cities and E-Participation that conceptualize the importance of this work. Section 3 presents theoretical reference and related work regarding to the techniques used in our work. Section 4 presents our proposed method. Section 5 describes a case study, considering (i) the collected data; (ii) its analysis; and (iii) opinions collected from software developers regarding to the steps of the presented method, implemented in the tool. Finally, Sect. 6 draws our final conclusions and discuss future work.

2 Smart Cities and E-Participation

Gil-Garcia, Pardo and Nam [7] affirms that, when considering the smartness of a city, rather than holding the dichotomy in terms of "being or not being" smart, smart city concept should hold a continuum in which local government, citizens and other stakeholders think about implementing initiatives that turns the city "smarter". In this way, they present a comprehensive view of smart city components and its elements, after compiling many different definitions found in literature and tools for evaluating or assessing the smartness of cities. Data management and information processing are two elements of the data and information component. In addition, [8] states that the smart city-building initiative "seeks to improve urban performance by using data, information and information technologies to provide more efficient services to citizens, monitor and optimize existing infrastructure, the collaboration between different economic actors and encourage innovative business models in both the public and private sector". In this scenario, it is important to consider the huge amount of data on social networks that is not commonly used by governors for firstly understand what issues in a city can emerge from this data source.

Also in this scenario, models raised in literature for e-participation, i.e., models for citizen interaction by e-government. Reddick, Chatfield and Ojo [3] choose and summarize three models, which can be represented on a continuum, being the managerial model the lowest form of e-participation, and the participatory model being the highest form of citizen interaction with government. In the managerial model, citizens are viewed as "costumers", and government provides information and services to satisfy the demand of these costumers, i.e., governments merely respond to their demands. In the consultative model, instead of being focused only on providing more efficient service delivery, the role for government is creating better policy decisions considering citizens claims and other inputs in the decision-making process. Finally, in the participatory model, there is a complex ow of information between governments and citizens, designed to enhance and shape policy. Regarding to the managerial model, it is worth to observe that several cities around the world provide open data in forms of reports, so that citizens can follow the actions of the government. The consultative model is somewhere in between the managerial and the participatory in its level of active. However, Reddick, Chatfield and Ojo [3] state that, when considering social networks for extracting knowledge about citizen claims, these different stages do not occur in a linear fashion. So, they propose a framework considering the use of text mining, analytics and visualization for explain their loop for e-participation. Specifically, visualization is a very important instrument for helping humans making sense of data [9]. This work specializes two aspects of this framework, when considering specific techniques for linking data, constructing visualization and presenting frequently addressed topics in comments on Social Network pages from local government agencies.

It is important to differentiate the terms "information" and "knowledge". Information is not a synonym for knowledge, which is an intellectual concept, referring to the condition of knowing or understanding something. Knowledge is organized information in people's heads. Selecting and analyzing data, information can be

produced; by selecting and combining information, knowledge can be generated; from this decisions can be made and action taken [10]. Data and Text Mining has been applied both for extracting information and knowledge from data [11]. In this work, we focus on using techniques commonly used in text mining for chunking and POS tagging words, brie y described next, from texts and data visualization techniques, aiming to enrich the comments and visualizing them (information extraction) to allow knowledge extraction by humans.

3 Theoretical Reference

3.1 Text Mining (TM)

Inzalkar and Sharma [12] state that the amount of unstructured stored data has tremendously increased, mainly in social networks. TM is the process of extracting interesting information or patterns from unstructured text from different sources. Tools for the different steps TM have the ability to analyze large amounts of text in natural language, and detect lexical and linguistic usage patterns, in an attempt to extract useful information [11, 12]. Nowadays, researchers use these tools in real-world applications, mining social networks for health or financial information, for identifying emotions about products and services, for example [13]. Also, sentiment analysis on social networks around political dimensions has been explored [14, 15].

Many of the NLP tasks involve searching for patterns in text that can be arduous if applying basic string operations. Regular expressions (`Regexes`) are primarily used in string search and substitution tasks in texts search and editing. `Regexes` are strings that define a search pattern [16]. To define a grammar, along with the use of regular expressions, it is necessary to use a technique called chunking [17], a fundamental mechanism of language: words can combine with other words, forming chunks. These can be combined with other chunks, to form even larger chunks, until a sentence is established. One way to construct the referred grammar is using POS (part-of-speech) taggers, which aims to assign a tag to each word in a text, or equivalently classify each word in a text to some specified classes such as norm, verb, adjective, etc. Both POS tagging and chunking are used in this work for extracting frequent topics from the comments.

3.2 Geocoding Data and Visualization

Up to our knowledge, there is not any work that links a comment/post in social networks to the location it mentions in its content. So, in this section we present similar works that reinforces the importance of our proposal. Cammarano et al. [18] considered the problem of visualizing heterogeneous data sets, describing a system capable of automatically finding specific information in the set, necessary to create a visualization. The researchers introduced a mechanism capable of describing views, regardless of the data, and a data recovery algorithm appropriate for a given view. Initial experiments demonstrated that the created system had the ability to find appropriate data through visualization. So, exploring the linkage and processing of different types of data to

generate different views can lead to enrich knowledge acquisition from unstructured data. This fact reinforces the importance of our method, as each comment must be geocoded by the zip code of some mentioned location, for generating the map visualization.

MacEachren, Brewer and Pickle [19] present a web analytics approach based on geovisualization using the social network Twitter, for supporting crisis management. Crisis management is the process by which an organization handles an unexpected event that threatens to undermine the organization and its audience. The proposed approach is implemented as a web application on a geographic map, which allows the user to search for information using indexing and tweet viewing, based on specific place and time characteristics. Our work is similar to this one, but different techniques for linking the data to be visualized had to be used. This is due to our interest in geocoding comments by its content, and not by the place it was posted.

Azevedo et al. [5] present an approach that enables the integration of unstructured data located in different public organizations. They present concepts and technologies that provide information visualizations from open data using Geographic Information Systems (GIS). From the proposed framework, the application is able to identify vulnerable communities and provide effective preventive and emergency actions. The main contribution of the work is to include the use of tools and methods for data publication. To validate their approach, the authors used flood data from the Rio Doce basin. Our work is similar to this one. However, we worked with the reality that comments on Facebook cannot be used as linked data, and neither the comments are geocoded by its content. So, we had to propose and implement another ways for recovering the geospatial data, as well as linking the comments to geospatial information.

Li et al. [6] present a survey about spatial technology and social media in remote sensing. They observe that there is a massive amount of data originated from remote sensing, social media, and GIS systems that are completely different sources of data. According to them, although important progress has been made in mining spatial and temporal data from social media, there is a need for investigating how these data can be used for decision making, particularly in the context of its integration with Geographic Information Systems (GIS). On the other hand, in our point of view, although there are many evolving solutions for GIS and possibilities for linking data, there are many data sources that do not follow this approach, as in our case. Also, [6] shows many applications, including in Brazil, that georeferenced image data and Twitter data, or similar approaches, for finding new knowledge in many specific domains. However, they do not present any work that explores generating new knowledge from different topics for understanding problems in a city, as we propose in this work.

4 Our Method

Considering all works described before, we understand that, up to our knowledge, there are gaps in literature that we explore in our method: (i) there is not any work that geocode comments based on some place mentioned in its content, which can help governments and citizens to visualize the distribution of comments over a city area;

(ii) there is not any work that extract frequent topics from comments using POS tagging and chunking, which can help governments and citizens to visualize what are the most frequent topics mentioned in some social network page.

Figure 1 shows a schematic representation of our proposed method in BPMN (Business Process Modeling Notation). Each of the activities are described in what follows:

1. Collecting Social Data: Extract all posts and comments in a social network page of a local government agency, in a given period.
2. Collecting Postal Service Data: Extract data from some postal service of a given city, containing the names of the streets, avenues and others; and, for each one, the respective neighborhood and the zip code.
3. Processing Comments: This is the core of our method, better explained later, responsible for identifying (i) the locations mentioned in comments; and (ii) their topics. We explain better this task later.
4. Calculating frequent topics: Detect the most frequent topics in the set of topics, counting the frequency of all the detected topics.
5. Geocoding locations (streets in general): Geocode each comment by the mentioned place, if it is founded.
6. Generating data visualization: Generate a data visualization in map format with the geocoded comments, and a list of the most frequent topics (in descending order of the number of mentions). The view should display, beyond the map, the city data (name and number of neighborhoods, sites, pages, posts and comments) and the most frequent topics, followed by the number of mentions.

Activities in Processing Comments: For this task, a tagger must be used for classifying a word according to its class. The word classes are noun, proper noun, personal pronoun, adjective, adverb, verb, numeral, preposition, subordinating and coordinating conjunction, interjection, and others. It is worth to mention that the tagger is dependent on the language of the text in analysis.

For identifying the topics, the following activities are executed for each comment: (i) Cleaning the comments: Remove abbreviation, links and words with more than one repeated letter; replace a repetition of more than one of the same punctuation symbol by only one symbol; and remove blank spaces before a punctuation symbol; (ii) Processing the comments (POS tagger): Divide each comment into sentences, and tag (each word of) each sentence using the given tagger; and (iii) Applying the chunking process: Process each tagged sentence by a syntactical analyzer, using a grammar with one regular expression (`RegEx` grammar) of the form:

```
<N.*>+<PREP.*>*<KS>*<ADJ>*<N.*>+
```

This analyzer looks only for subjects that form a sequence of a noun (N, which may be a regular or a proper noun), followed by a preposition, a subordinating conjunction, adjective or another (regular or proper) noun. The symbol "*" means that sequences that attends this grammar has no limit of size. The output is a collection of subjects

found in all sentences of the comment. The output of this task is a set with all topics found in all the set of considered comments.

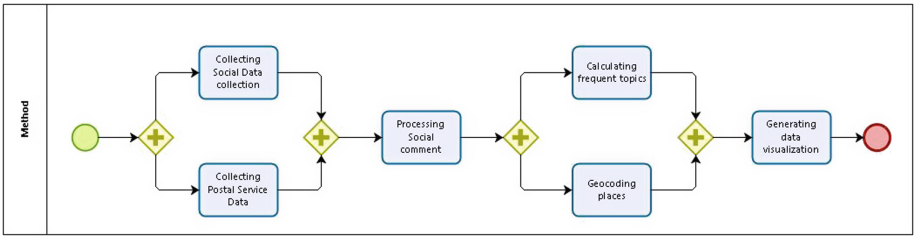

Fig. 1. Our method for processing comments in local government pages in social networks for knowledge acquisition.

For identifying the locations mentioned in comments, the following steps are executed for each comment: (i) Cleaning the comments: The same in the previous task, added to another activity: Remove stopwords (prepositions, pronouns and others); (ii) Processing the comments: Look for words that represent the type of the location (street, avenue, plaza and others); and (iii) Retrieving the names of possible locations: Retrieve from the database all possible locations of the type found in the previous activity, and find which one occurs in the comment. Return the related zip code of the location found.

5 A Case Study

In order to validate our method, we implemented a web system[1]. It was logically divided into three modules: extraction (responsible for Activities 1 and 2), processing (responsible for Activities 3, 4 and 5) and visualization (responsible for Activity 6). All of them were implemented using.Net Core, specifically ASP.NET for constructing web applications, and Angular and Bootstrap frameworks for the graphical interface. We also used NLTK[2] (Natural Language Toolkit) for facilitating the POS Tagging and Chunking activities. We also constructed a tagger using a corpus for Brazilian Portuguese called Mac-Morpho[3] [20].

5.1 Collecting and Processing the Data

For the analysis of the operation of the tool, several information from different sources were collected, which are (i) the city's streets and neighborhoods; (ii) address geocoding; and (iii) comments of interest in the social network. Our case study

[1] Available at github.com/pedroccrl/tcc.

[2] Available at https://www.nltk.org/.

[3] Available at http://nilc.icmc.usp.br/macmorpho/.

addressed the municipality of Rio das Ostras, in the state of Rio de Janeiro, where a part of our research group is located. In what follows, we detail the data and information collection for each of the three items:

(i) Rio das Ostras streets and neighborhoods: We extracted the information from the streets and districts of the city of Rio das Ostras on the website of the Brazilian Post Office[4]. It was difficult to obtain this information, because it is necessary to specify at least the neighborhood to retrieve the information of a municipality's public place. We obtained this information (neighborhoods of the city) with an employee of the Information Technology Department of Rio das Ostras. Also, to increase the efficiency of the application and to be able to be used in other contexts, we developed a robot to go through the Brazilian post office and obtain the data required. This collected data was inserted into the database for each neighborhood. A total of 1505 names of locations (streets, avenues, and so on) were obtained by the robot.

(ii) Address Geocoding: We used Google Maps Geocoding API service for geocoding the addresses (names of the streets), associating coordinates with each zip code. All the 1505 locations were geocoded through this process.

(iii) Frequent Topics in Comments: Facebook Graph API was used for collecting Facebook data. Each request in this API returns a document. For this, an Access Token, or authentication key, is needed. For this, we created an application in the Facebook Graph page. We collected posts and comments from the Facebook page "Cidadão Riostrense" (in Portuguese, which could be freely translated to Citizens of Rio das Ostras). We collected a total of 881 posts and 32157 comments between January 2016 and November 2017.

5.2 Visualizing the Data

Processing the collected data, we identified the mention of locations in 2054 comments, while in the georeferenced sites, we found 184 comments. Figure 2 shows the map present on the screen of the constructed web system, with the extracted information from the collected data. In the original web system page, the left column displays the city data (name and number of neighborhoods – 66, locations – 1505, pages – 1, posts – 881, and comments – 32157), the middle one displays the refereed map, and the right column displays the most frequent topics, followed by the number of mentions. From the list of the hundred most frequent topics, we filtered the most twelve important ones, shown in Table 1. We observed that the main topic in the collected comments was 'street lighting', and this is not an information that could be obtained in any other place. There are some other curious observations in this list. The mayor's name ("Carlos Augusto") and "Public Power" appeared many times as topics, and typically was related to some negative sentiments, when we observed the entire comments. Also interesting is the topic "God Comfort". This indicates that sentiment analysis in comments of local government agencies should also be interesting. A deeper

[4] We used the search tool available at http://www.buscacep.correios.com.br/sistemas/buscacep/resultadoBuscaLogBairro.cfm.

investigation of the remaining topics should be interesting with the local government, but it should not be executed up to the publication of this work.

Fig. 2. Map of the constructed website for Rio das Ostras city.

Table 1. Frequent topics and their frequencies. "Carlos Augusto" is the mayor's name.

Topic	Frequency	Topic	Frequency
Public lighting rate	181	Municipal guard	41
Carlos Augusto	76	Lack of education	40
Public Power	75	God comfort	39
Fireworks	51	Public money	36
Basic sanitation	46	Garbage collection	34

It is important to observe that manually processing more than 30 thousand comments for extracting comments is not an easy task. When the task is complete by humans, probably the information will be obsolete. A more difficult task would be manually identifying the streets (locations), present in the content of the comments, and also geocoding them. This task was only possible to be executed by the computer using APIs and services available in the Web, as discussed before. So, our method allowed us to observe which citizens' claims emerge in a particular city, helping us to better understand people's real feelings, and which locations have the most problems.

5.3 Evaluating the Usefulness of the Tool

In order to try to identify in which level people graduated in computing undergraduate courses can use the tool for obtaining new knowledge, and how the tool can be evolved, we conducted an experiment with four participants, aiming to obtain a qualitative insight in the usefulness of the tool. For this, we constructed a question and answer form, containing five questions. Questions Q.1 to Q.4 were of the Likert scale

type, and Q.5 asked the user to indicate how the tool could be evolved. The first four questions are: [Q.1] What level of difficulty did you feel in using the tool to identify which regions of Rio das Ostras are most commented? [Q.2] How much difficulty did you feel in using the tool to identify the most frequent topics? [Q.3] What level of difficulty did you feel in acquiring knowledge about regions with the most problems in the municipality of Rio das Ostras? [Q.4] What level of difficulty did you feel in acquiring knowledge about the most frequent topics in the comments? Q.1 and Q.2 scales range from 1 to 5, where 1 means "Very difficult" and 5 means "Very easy". Q.3 and Q.4 scales range from 1 to 10, where 1 means "I could not extract new knowledge" and 10 means "I discovered many new and interesting things". The difference is due to the understanding that the question of extracting new knowledge can be much more sensitive when the user changes. Tables 2 and 3 show the number of responses obtained by scaling (first column) for each question (remaining columns). Table 4 shows the answers of the participants in Q.5. It is worth to notice that we chose software developers for this first analysis due to their ability to understand how is difficult to analyze unstructured data. They suggested interesting ways to evaluate our tool.

Table 2. Answers from participants to Q.1 to Q.4.

Q.	1	2	3	4	5
1	0	0	0	2	2
2	0	0	1	2	1

Table 3. Answers from participants to Q.1 to Q.4.

Q.	1–6	7	8	9	10
3	0	1	1	0	2
4	0	0	2	0	2

Table 4. Answers to question Q.5

Q.5: How do you believe that the tool could be evolved?	
1.	"Very good idea and tool! On evolution: to make comparative analyzes in the future on the evolution of topics/comments, to identify if there was any progress in the problems and what new ones arose, allowing analyzes on the performance public agencies and perception of the population in the region"
2.	"[The system] could have a space to register the responsible public agencies concerned [to the related problems], and a notification for submission and possibility of response of the type "complaint here""
3.	"Twitter API can be integrated, which allows users to get more information about the problems (so other channels can take advantage of this feature)"
4.	"In the technical part, I think that an implementation using Artificial Intelligence over the topics found is required. Determining the really important topics is worth – unnecessary topics are appearing"

6 Conclusions and Future Work

This work presents a method that allows extracting new information in comments in a social network page of a local government agency, through identifying the most frequent topics and geocoding these comments based on the location it refers to. Our initial experiments allowed us to observe that visualizing the map with the geocoded comments, joined to the frequent topics, allowed (advanced) users extracting knowledge from the collected comments.

Based on literature review in e-participation and smart cities, citizens' claims are information of great value to the government. In this work, we used two forms of analyzing and extracting knowledge about their claims posted in social networks, which are frequent topics and geovisualization. We understand that this is the first step for exploring this huge amount of data. In this way, sentiment analysis and sarcasm identification others are interesting methods for being investigated in this scenario. Specifically, this is a challenge in Portuguese language. In the future, we intend to investigate the use of these approaches to measure and interpret citizen claims in social network.

We also consider the following limitations of our work, for improvement in the future: (i) We did not evaluate the precision of our method from Information Retrieval perspective. For this, we need to collect a sample of the comments, label the comments with the name of the location (street and others), use the technique for detecting the name of the location, and verify the precision of the technique. This type of information, if precise, can be added as a metadata of the comments; (ii) Other ways of geocoding comments, as using, for instance, geolocation of the user that created the comment, could be interesting, in the case of this information is available; (iii) Our qualitative experiment allowed to observe future improvements of the method and the tool, although the number of participants is too low. We desire to evaluate our tool using citizens of different profiles, as well as investigate the usefulness of the toll with people working with local government agencies.

Acknowledgements. We would like to thank the precious comments of the anonymous referees that helped us to improve our work. We also thank CAPES, connected to the Brazilian Ministry of Education, FAPERJ (Foundation of Support to the Research of State of Rio de Janeiro) and CNPq (Brazilian National Council for Research) for partially funding this work.

References

1. Terán, L., Kaskina, A., Meier, A.: Maturity model for cognitive cities. In: Portmann, E., Finger, M. (eds.) Towards Cognitive Cities. SSDC, vol. 63, pp. 37–59. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-33798-2_3
2. Dameri, R.P., Rosenthal-Sabroux, C. (eds.): Smart City: How to Create Public and Economic Value with High Technology in Urban Space. PI. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06160-3

3. Reddick, C.G., Chatfield, A.T., Ojo, A.: A social media text analytics framework for double-loop learning for citizen-centric public services: a case study of a local government Facebook use. Gov. Inf. Q. **34**(1), 110–125 (2017)
4. Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., Trevisan, D.: Transparency in practice: using visualization to enhance the interpretability of open data. In: Proceedings of 18th International Annual Conference on Digital Government Research, pp. 139–148. ACM, New York (2017)
5. Azevedo, P.C.N., Pinto, V.A., Bastos, G.S., Parreiras, F.S.: Using linked open data in geographical information systems. In: Grueau, C., Gustavo Rocha, J. (eds.) GISTAM 2015. CCIS, vol. 582, pp. 152–166. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29589-3_10
6. Li, J., Benediktsson, J.A., Zhang, B., Yang, T., Plaza, A.: Spatial technology and social media in remote sensing: a survey. Proc. IEEE **105**(10), 1855–1864 (2017)
7. Gil-Garcia, J.R., Pardo, T., Nam, T.: What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization. Inf. Polity **20**(1), 61–87 (2015)
8. Marsal-Llacuna, M.L., Lopez-Ibañez, M.B.: Smart urban planning: designing urban land use from urban time use. J. Urban Technol. **21**(1), 39–56 (2014)
9. Baker, J., Jones, D.R., Burkman, J.: Using visual representations of data to enhance sensemaking in data exploration tasks. J. Assoc. Inf. Syst. **10**(7), 533–559 (2009)
10. Webb, S.P.: Knowledge management: Linchpin of change. Routledge, London (2017)
11. Aggarwal, C.C., Zhai, C.X. (eds.): Mining Text Data. Springer, New York (2012). https://doi.org/10.1007/978-1-4614-3223-4
12. Inzalkar, S., Sharma, J.: A survey on text mining-techniques and application. Int. J. Res. Sci. Eng. **24**, 1–14 (2015)
13. Hirschberg, J., Manning, C.D.: Advances in natural language processing. Science **349** (6245), 261–266 (2015)
14. Weitzel, L., Bernardini, F., Quaresma, P., Alves, C.A., Zacharski, W., de Figueiredo, L.G.: Brazilian social mood: the political dimension of emotion. In: Fuhr, N., et al. (eds.) CLEF 2016. LNCS, vol. 9822, pp. 247–252. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44564-9_23
15. Paula Filho, W., Rosseti, I., Viterbo, J.: On tweets, retweets, hashtags and user profiles in the 2016 american presidential election scene. In: Proceedings of 18th International Annual Conference on Digital Government Research, pp. 120–128. ACM, New York (2017)
16. Chapman, C., Wang, P., Stolee, K.T.: Exploring regular expression comprehension. In: Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering, pp. 405–416. IEEE Press, Piscataway (2017)
17. Lu, Q., Xu, C., Liu, H.: Can chunking reduce syntactic complexity of natural languages? Complexity **21**(S2), 33–41 (2016)
18. Cammarano, M., et al.: Visualization of heterogeneous data. IEEE Trans. Vis. Comput. Graph. **13**(6), 1200–1207 (2007)
19. MacEachren, A.M., et al.: Geo-twitter analytics: applications in crisis management. In: Proceedings of the 25th International Cartographic Conference, pp. 3–8. GeoVISTA Publications, State College (2011)
20. Fonseca, E., Aluísio, S.M., Rosa, J.: Evaluating word embeddings and a revised corpus for part-of-speech tagging in Portuguese. J. Braz. Comput. Soc. **21**(2), 1–14 (2015)

Author Index