

Collaborative Practices and Multidisciplinary Research: The Dialogue Between Entrepreneurship, Management, and Data Science



**Riccardo Fini, Monica Bartolini, Stefano Benigni, Paolo Ciancarini,
Angelo Di Iorio, Alan Johnson, Marcello Maria Mariani, Silvio Peroni,
Francesco Poggi, Einar Rasmussen, Riccardo Silvi, Maurizio Sobrero,
and Laura Toschi**

Abstract Digital technologies and their applications are systematically altering established practices and making new ones emerge in different realms of society. Research in social sciences in general and management in particular is no exception, and several examples that span various fields are coming into the spotlight not only from scholarly communities but also the popular press. In this chapter, we focus on how management and entrepreneurship research can benefit from ICT technologies and data science protocols. First, we discuss recent trends in management and data science research to identify some commonalities. Second, we combine both perspectives and present some practical examples arising from several collaborative projects that address university–industry collaborations, the impact of technology-based activities, the measurement of scientific productivity, performance measurement, and business analytics. Implications for using data science in entrepreneurship and management research are discussed.

R. Fini (✉) · M. Bartolini · R. Silvi · M. Sobrero · L. Toschi
Department of Management, University of Bologna, Bologna, Italy
e-mail: riccardo.fini@unibo.it

S. Benigni
Imperial College Business School, London, UK

P. Ciancarini · A. Di Iorio · S. Peroni · F. Poggi
Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

A. Johnson
Nord University Business School, Bodø, Norway

Ratio Institute, Stockholm, Sweden

M. M. Mariani
Henley Business School, University of Reading, Reading, UK

E. Rasmussen
Nord University Business School, Bodø, Norway

Keywords Entrepreneurship · Data science · University-industry collaborations · Technology innovation management · Scientometrics · Strategy · Business performance analytics

1 Introduction

Huge amounts of data (“*Big Data*”) are produced inside and outside contemporary companies by people, products, and business infrastructures. However, it is often difficult to know how to transform these data flows into effective strategies and actionable plans. Data science¹ has potential for companies of all types to find patterns and models in these data flows and use them as the basis for disruptive analyses and derived software platforms.

From Radio-Frequency Identification sensor data to customer loyalty programs, predictive analytics can improve customers’ engagement and companies’ operational efficiency. Indeed, several precious insights await organizations that can exploit findings obtained from data science. Data science is a novel discipline, which can enable any effort of digital transformation. Hence, digital transformation, being defined as ‘the acceleration of business activities, processes, competencies, and models to fully leverage the changes and opportunities of digital technologies and their impact in a strategic and prioritized way’ (www.i-scoop.eu), concerns the need for companies to enact digital disruption and remain competitive in an ever-changing competitive environment.

*Big Data*² is generated continuously, both inside and outside the Internet. Every digital process and economic transaction produces data, sometimes in large quantities. Sensors, computers, and mobile devices transmit data. Much of this data is conveyed in an unstructured form, making it difficult to put into database tables with rows and columns. Aiming at searching and finding relevant patterns in this complex environment, data science projects often rely on predictive analytics, involving ‘machine learning’³ and ‘natural language processing’⁴ (NLP), as well as on

¹Data science is defined as ‘a set of fundamental principles that support and guide the principled extraction of information and knowledge from data’ (Provost and Fawcett 2013, p. 52).

²*Big Data* can be defined as ‘the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value’ (De Mauro et al. 2016).

³Technology now makes it possible for software solutions to learn and evolve. Software with machine learning capabilities can produce different results given the same set of data inputs at different points in time, with a learning phase in between. This is a major change from following strictly static program instructions, like most of the artificial intelligence models from the 1990s.

⁴Technology now makes it possible for software solutions to talk and interpret language from humans, be it in speech or in documents. Software with semantic processing ability is able, for instance, to perform sentiment analysis, a kind of analytics able to scan large corpora of documents to determine the polarity about specific entities or concepts. It is especially useful for identifying trends of opinion in a community, or for the purpose of marketing.

cloud-based applications.⁵ Computers running machine learning or NLP algorithms can explore the available information by sifting through the noise created by *Big Data's* massive volume, variety, and velocity.

The societal impacts of these changes are being debated daily (*New York Times International*, March 1st 2017) and the amount of evidence produced to stress how 'things will never be the same' combines easy-to-communicate anecdotal evidence and more rigorous analyses. The research community is certainly among the various fields where the impact of machine learning, NPL, and cloud architectures is redefining the rules of the game. While clearly relevant in many computationally intensive and data-dependent research endeavours, new opportunities are also opening for unexplored alternatives in other research domains where classification, parsing, and clustering of text and images have so far depended mostly on human-centered activities and interpretation. Management and entrepreneurship research are no exception on several grounds.

First, the way managerial and entrepreneurial activities in companies and institutions are being affected by these changes is clearly an area of increasing interest. In a recent book collecting evidence of several years of research, for example, Parker et al. (2016) analyse how two-sided network effects can be leveraged to build effective cloud-based product platforms, showing how data-driven technologies can be key determinants of competitive advantage. Arun Sundararajan (2016) reached a similar conclusion in his extensive analysis of the different forms of sharing economy and their dependence on several enabling factors all related to the similar evolutions and patterns in data.

Second, the opportunities embedded in the new technologies and methods for data gathering and analyses are being explored to improve both efficiency and effectiveness of sample collections, and to design original alternatives to collect and manipulate empirical evidence. In a recent editorial published in the *Academy of Management Journal*, George et al. (2016) discuss at length how to frame the challenges faced. More precisely, they suggest distinguishing between the effects in management research from data collection, data storage, data processing, data analysis, and data reporting and visualization. Like in many social sciences, whenever research questions are related to specific occurrences, any opportunity to extract, accumulate, and analyse multiple episodes and instances helps to develop hypothesis testing and to identify patterns and regularities. The power of data science goes well beyond the contributions offered by large databases, which have significantly changed the field since the early nineties. However, these new opportunities are still far from being incorporated into doctoral programs for the new generations

⁵Companies can take advantage of the elastic nature of the cloud and deploy their products by exploiting the flexibility, agility, and affordability provided by cloud platforms. Cloud-based applications provide global support and real-time access to *Big Data* from anywhere in the world at any time. By replicating the environment, multiple enterprise environments remain in sync, and their data flows can be easily integrated. Because applications in the cloud are always deployable, always available, and highly scalable, continuous, agile innovation becomes an objective achievable by any business.

of researchers, and they certainly require the education of many editors to be able to properly staff their reviewing teams to ensure that they are adequately equipped to evaluate the pros and cons of applications of new methodologies that leverage data science advances.

Finally, major changes in decision-making processes touch the fundamental bases of several theories and conceptual frameworks. From the notion of bounded rationality (Simon 1972), to the interplay between local and distant search (March and Simon 1958), to the impact of information asymmetry reduction opportunities to determine governance structure (Nayyar 1990), scholars of management and entrepreneurship are witnessing an unprecedented impact of technologies, not simply on practices and methods, but on constructs and theories as well. Take transaction costs economics, for example, introduced by the Nobel laureate Oliver Williamson (1979), and consider a reinterpretation of the continuum between markets and hierarchies under the currently plummeting cost and time needed to gather and analyse the necessary information. Opportunistic behaviours can be thus anticipated with greater precision thanks to more efficient simulations based on evidence recovered from various and widespread sources such as news, blogs, or interactions on social networks. Furthermore, in the context of credit scoring for trading partners, the traditional reference of the so-called FICO score,⁶ provided by reputable intermediaries—who parse through dedicated sets of private information retained by various financial institutions—is being challenged using algorithms to determine organizations' risk profiles based on their relationships and positioning in multiple social networks.

We believe we are only at the beginning of an exciting time full of unexplored opportunities worth pursuing within and across disciplines. Hence, the aim of this chapter is to explore how entrepreneurship and management research can collaborate with data science to benefit from new digital data opportunities. To do so, we present five case examples, elaborating on how researchers can make use of data science in different areas of management research. The next section provides an overview of the selected examples, before each case is presented in more detail in the following sections. The chapter concludes by offering some remarks and implications for further research in the area of entrepreneurship.

2 Overview of Cases

To advance our understanding of the new data possibilities, we explore some preliminary ideas originating within five different collaborative projects, operating at the interface between management and data science research. These cases have been selected to represent a variety of examples of potential and ongoing research

⁶First introduced in 1989 by FICO, a public company established in 1956 as Fair, Isaac, and Company.

that can both inspire and provide specific advice to management and entrepreneurship scholars on how to seize new opportunities in an increasingly digitalized world. First, we look at the case of collaboration between entrepreneurial firms and universities and how data science techniques could be applied to shed light on processes that are largely unknown at present. The recent advent of remote sensing, mobile technologies, novel transaction systems, and high-performance computing offers opportunities to understand trends, behaviours, and actions in a manner that was not previously possible.

Second, we investigate the case of technology innovation management and the challenge of measuring the impact of technology-based activities. A common indicator is the patent protection of intellectual property rights, which is often based on relations between variables at different levels of analysis, using data that is uncodified, dynamic, and generally unavailable in a single dataset. The field of semantic technologies can offer key complementarities to support the (semi-) automated creation of structured data from non-structured content and generate meaningful interlinks.

Third, we explore the case of measuring scientific productivity, which is at the heart of scientometrics approaches. Measures of scientific constructs using data science techniques are subject to the same reliability and validity concerns as any other source of measurement (e.g., questionnaire responses, archival sources), where researchers struggle with the balance between the theoretical concepts they are interested in (e.g., scientific progress), and the empirical indicators they are using to operationalize them (e.g., publications and citations). In scientometrics, measures largely emerge from how publication practices are recorded, and how these archival records represent intentional individual or collective strategies and outputs.

Fourth, we present a case combining entrepreneurship and strategic management interests in the tourism and hospitality industries. In particular, a large amount of unstructured data, such as online searches, accommodation bookings, discussions, and image and video sharing on social media produced by tourists and companies, as well as online reviews, has profoundly affected the whole value chain of different economic agents in the field. And yet, a vast number of destinations as well as SMEs often ignore or underuse this type of data because it is unstructured and therefore difficult to analyse and interpret. Several applications, developed to solve different problems, could offer viable opportunities to overcome these limitations and strengthen local economic systems.

Fifth, our last case takes the collaborations between management and ICT one step further. Specifically, it explores the role of business performance analytics as a valuable support tool for management-related issues by transforming data into information valuable for decision-making. It focuses on the strategic relations occurring between the two domains and their effect on the abilities to collect, select, manage, and interpret data to generate new value.

2.1 *Five Examples of Cross-Fertilizations Between Management, Entrepreneurship, and ICT*

Case 1: University–Industry Collaborations

University–industry collaboration (UIC) refers to the interaction between industry and any part of the higher educational system, and is aimed at fostering innovation in the economy by facilitating the flow of technology-related knowledge across sectors (Perkmann et al. 2011). Of late, there has been a substantial increase in UICs worldwide and numerous studies that investigate questions in the field.

Based on a systematic review of the literature, Ankrah and AL-Tabbaa (2015) propose a conceptual framework highlighting five key areas of the literature on UIC that required further investigation. First, currently employed measures to evaluate outcomes of collaboration are essentially subjective and more objective measures of the effectiveness of UIC need to be explored. Second, more research is needed to examine the boundaries of the role of government in UICs within the Triple-Helix model (Etzkowitz and Leydesdorff 2000). Third, there is a need to conduct comparative studies across different countries in relation to UIC. Fourth, most of the studies found in the literature are cross-sectional and a longitudinal line of research is needed to explore cause-effect relations in the evolution of UICs. Finally, the impact of academic engagement as a form of UIC on the outcomes is almost completely overlooked. Accounts of both formal activities, such as contract research and consulting, and informal activities, such as providing ad hoc advice and networking with practitioners, are largely unexplored in the literature and could provide supporting evidence to an intangible potential value for UIC (see also Perkmann et al. 2015).

Among the research gaps in the literature listed above, informal inter-organizational ties offer a fruitful avenue for the application of recent developments in ICT and data science. One of the main outcomes of UIC, namely the exchange of knowledge and technology, occurs by means of formal and informal ties both at the individual and organizational levels. Formal links facilitate knowledge transfer while informal links generate knowledge creation (Powell et al. 1996). Notably, among the industrial partners, entrepreneurial firms rely significantly on informal, or embedded (Granovetter 1985), links during the early stages of their life cycle, when they most need to acquire and develop new knowledge and are most likely to engage with universities for this purpose (e.g., Anderson et al. 2010).

Informal ties remain largely unexplored within the context of UIC, as well as in the innovation and inter-organizational networks literature (West et al. 2006). Data science and ICT can now offer a great deal of new information or *Big Data* that can be leveraged to further explore the nature of informal links, the extent to which they permeate inter-organizational collaborations and their main antecedents and consequences. Informal network ties may be captured by exploiting the wealth of data stored and exchanged on social network sites (SNSs), making large-scale collection of high-resolution data related to human interactions and social behaviour economically viable. There is increasing evidence of entrepreneurs' growing use of

Facebook, LinkedIn, Instagram, Twitter, and other SNSs. These sites have the capacity to help entrepreneurs initiate weak ties (Morse et al. 2007) and manage strong ones (Sigfusson and Chetty 2013).

Virtual networking is complementary to real-world interactions and facilitates the establishment of new connections and the development of trust relationships. Therefore, even simple measures of social network interconnectedness between industry and university actors have the potential to uncover a great deal of existing informal ties and on-going informal collaborations. The data on SNS links are generally publicly available and can be collected by means of various web-scraping methods. Complementary data can be obtained with the aim of recently developed software tools. For instance, NVivo 11's tool can code Facebook screen shots, providing textual and visual data for the analysis of different kinds social interaction. Another example is the software CONDOR (MIT Center for Collective Intelligence) that can identify subnetworks of people talking about the same topics by sourcing various SNSs and applying clustering and sentiment analysis techniques.

Furthermore, the new data science methods and tools allow the UIC researchers to progress significantly to analyse not only the extent of the network of informal ties, but also the actual flows of information that occur through those channels. Large amounts of data and analytic gold lie hidden in multiple formats such as text posts, chat messages, video and audio files, account logs, navigation history data, profile biographic and meta- data, and other textual and visual sources. Email communications significantly extend the range of the sources from which this rich, high-granularity data can be pooled. This wealth of data can be mined using content analysis and machine learning techniques to measure the extent and nature of the information exchanged. It is possible, for example, to determine whether communications occur at the personal level, aimed at the development and maintenance of personal trust relationships; or at the technical level, aimed at the exchange of both tacit and explicit knowledge, the former being vital for the innovation process and overall UIC outcomes. In this regard, evidence suggests that virtual communication exchanges tend to shift from explicit, more codified knowledge at the beginning of the relationship towards tacit, more detailed knowledge exchange when the collaboration relationship matures (Hardwick et al. 2013). Nevertheless, Polanyi (2013) points out that the narrower channel of virtual communication may restrict the transfer of tacit knowledge and that this is best shared in face-to-face interactions.

Developing the tools to leverage the newly available streams of data can potentially answer these and several other questions related to UIC and offer great promise to both management scholars and policymakers. Should the newly available data reveal significant informal links between participants of successful collaborations, the operationalization practices of UIC might need to be extended to include processes and activities that incentivize the creation and development of informal networks. While these efforts are already made in practice (Ritter and Gemunden 2003), the insights provided by the analytical tools of data science might offer new, smarter ways to promote engagement in informal activities.

Therefore, we argue that ad-hoc data science models and tools to tap into the abundant wealth of data offered by newly available sources such as social media and

organizations' unstructured data offer great opportunities to deepen our understanding of inter-organizational networks and significantly boost the outcomes of UIC.

Case 2: Technology Innovation Management

Technology Innovation Management (TIM) refers to the study of the processes to launch and grow technology businesses and the related contingent factors that affect the opportunity for, and constraints on, innovation (Tidd 2001). Technology entrepreneurship, focused on the development and commercialization of technologies by small and medium-sized companies; open source business, analysing firms adopting a business model that encourages open collaboration; and economic development in a knowledge-based society (McPhee 2016) are some commonly investigated topics in this field.

The heterogeneity and complexity of this area is a fruitful field to show how artificial intelligence and web data may open important opportunities to foster research. Digitalization affects individual and team behaviours; organizational strategies, practices, and processes; industry dynamics; and competition. In the paper by Droll et al. (2017), for instance, a web search and analytics tool—the Gnowit Cognitive Insight Engine—is applied to evaluate the growth and competitive potential of new technology start-ups and existing firms in the newly emerging precision medicine sector.

More generally, empirical research in TIM is often based on relations among variables at different levels of analyses, whose data are uncodified, dynamic, and generally unavailable in a single dataset. Thus, providing a longitudinal and multilevel analysis is a crucial requirement for advancing research in TIM. A comprehensive data science approach, characterized by richness of data, allows researchers to answer new questions; avoid premature conclusions; identify fine-grained patterns, correlations, and trends; and shed new light on observed phenomena.

However, this goal poses two challenges: (i) automated importing and cleaning of data and (ii) dis-ambiguous integration of fragmented data. The first issue is a well-known aspect of the data science domain. When considering a large corpus of non-structured data that should be converted into structured information to address analytic and sense-making tasks, the use of automatic and/or semi-automatic tools is the best (and probably the only) way to complete the conversion in a reasonable timeframe. Several tools allow the automatic analysis—e.g., Apache UIMA (Ferrucci et al. 2009)—and conversion—e.g., DeepDive (Zhang 2015) and ContentMine (Arrow and Kasberger 2017)—of unstructured content; they are supported by large communities of computer scientists and data scientists to guarantee their sustainability and evolution over time. However, these tools represent only preliminary steps toward increasingly structured data automation processes.

In the past 15 years, web technologies have been radically expanded and now include several languages and data models that allow anyone to make available structured data on the most disruptive communication platform in recent decades—the web. These new tools, named *semantic web* technologies, enable researchers to describe structured data on the web by means of Resource Description Framework

(Cyganiak et al. 2014), share these data according to common vocabularies defined by using OWL (Motik et al. 2012), and query them by means of an SQL-like language called SPARQL (Harris and Seaborne 2013).

The real advantage of using such technologies is that the data are not enclosed in monolithic silos, which usually happens with common databases; rather, they are available on the web to anyone as a global and entangled network of linked resources. These resources can be browsed and processed by means of standard languages, and the statements they are involved in can be used to infer additional data automatically by means of appropriate mathematical tools. These semantic web technologies are the most appropriate mechanism to expose the structured data, obtained from a conversion of unstructured information, in a shared environment such as the Web, and for enriching them by adding new links to other relevant and even external data and resources that someone else may have made available with the same technologies.

The use of these technologies within the scholarly communication has resulted in a new stream of literature, *semantic publishing* (Shotton 2009). Broadly speaking, semantic publishing concerns the use of web and semantic web technologies and standards for enhancing scholarly and/or industrial work semantically (by means of RDF statements) so as to improve its discoverability, interactivity, openness, and (re) usability for both humans and machines. There are already examples of projects that have begun to make scholarly-related data available on the web by means of semantic web formats, such as OpenCitations (<http://opencitations.net>), which publishes citation data (Peroni et al. 2015), Open PHACTS (<https://www.openphacts.org/>), which makes available data about drugs (Williams et al. 2012), and Wikidata (<https://wikidata.org>), which contains encyclopaedic data (Vrandečić and Krötzsch 2014). However, as far as we know, these technologies have not been used yet for sharing and interlinking resources in several TIM contexts.

In the following, we will present two applications that highlight the power of data science in TIM projects. Specifically, the first example shows the use of disambiguation techniques to address problems of lack of unique identification names, derived by common errors of data entry, incorrect translations, abbreviations, name changes or mergers between institutions. In the second example, Natural Language Processing (NLP), which involves automatic processing by an electronic calculator of information written or spoken in a natural language, is applied to deconstruct data and import them into a final dataset.

The PATIRIS (Permanent Observatory on Patenting by Italian Universities and Public Research Institutes) project (<http://patiris.uibm.gov.it>) maps patent data over time with the aim to analyse the *innovative productivity* of Italian public research institutes. Rather than focusing on single patent documents, PATIRIS allows users to analyse patent groups—in different countries and over time—related to a common invention, defined as ‘patent families.’ The use of patent data to measure innovative activity requires precise arrangements to properly characterize inventions rather than single patent documents. The lack of unique IDs for patent assignees by the various international patent authorities generates a significant number of name variants, creating substantial distortions. For this reason, disambiguation techniques of the

assignee names are required to match a single institution to multiple variants of its name. This problem may be addressed manually with a limited number of observations but automated and structured ICT techniques are recommended for larger samples. This is also particularly useful when an update of the data over time or integration of information from different data sources are required. PATIRIS, for instance, updates its data twice per year and obtains assignee-level information through the MIUR (Ministry of Education, Universities and Research) dataset (www.miur.it).

The TASTE (TAKing STock: External engagement by academics) project (<http://project-taste.eu>) has the aim to systematically map academic entrepreneurship from Italian universities and better understand the determinants and consequences of science-based entrepreneurship (i.e., Fini and Toschi 2016; Fini et al. 2017). Key distinguishing features of the project include (i) the adoption of a multi-level approach, (ii) the integration of multiple data sources, and (iii) the longitudinal structure of the data. TASTE integrates five different domains at the individual-, knowledge-, firm-, institutional- and contextual-level. More precisely, it analyses about 60,000 academics, their 1000 patents, and 1100 spin-offs, characterizing their 95 universities and 20 regions for the period 2000–2014. To obtain such a multilevel structure, the researchers integrated data from ad-hoc surveys sent to university research offices, technology transfer offices, spin-offs and entrepreneurs; LinkedIn; the European Patent Office and PATIRIS; the Italian Ministry of the University and Research; Eurostat; and others. In this research design, the automated and structured retrieval, which was designed to import, clean, and integrate the data, was critical for the integrity of the data and the feasibility of the project. Recently, the project has also implemented *semantic publishing* techniques.

These examples show how the combination of ICT and management research techniques allow academics to investigate new and unexplored research questions (George et al. 2016) by exploiting the three core characteristics of big data: ‘big size’ of datasets, ‘velocity’ in data collection, and ‘variety’ of data sources integrated in a comprehensive way (McAfee and Brynjolfsson 2012; Zikopoulos and Eaton 2011).

Case 3: Scientometrics

Scientometrics is a multi-disciplinary field that aims at studying ways for measuring and analysing progress in science and related technologies through various approaches. But who decides what constitutes *scientific progress* and whether specific people, places, and times have helped science to progress or not, and on what basis? What are the criteria for researchers and professors to be promoted? What are the criteria for whether academic departments continue or get cut, and whether research projects get funded or not? There is an increasing trend in western countries towards using ‘objective’ criteria to make such decisions but the scare quotes indicate that these criteria are at least partially open to strategic manipulation and potentially outright gaming. We discuss some of these dangers and potential strategies to ameliorate them below.

Operational classifications in social science are called *coding*: when social scientists assess an observation into a specific class (progressive/not progressive) or

assign it a specific number (i.e., a score of five, as opposed to four on a clearly articulated anchoring scale). Accumulated publications and citations, corrected for self-citation, weighted for number of co-authors, and aggregated across individuals, departments, faculties, and institutions, is an example of a coding process. Coding is a fundamental part of the process of measurement. We expect scientists to design appropriate measures and to implement them faithfully during data collection. Properly defined and executed measurements provide us with a precise picture of the way things are (e.g., scientific progress) that we want to study and give us the basic information for our scientific generalizations and probabilistic models (Cartwright 2014), which we may use to change the world around us and design interventions in that world where necessary.

Measurement is finding a grounded and systematic way to assign values or numbers to observations (i.e., putting them into categories in a rule-governed and consistent way). Measurement involves three steps that are interrelated, which should not only be consistent but also mutually supporting (Cartwright and Runhardt 2014):

1. *Characterisation*: lay out clearly and explicitly what the quantity or category is, including specific features of it for researchers to make use of when assigning numbers or categories to observations.
2. *Representation*: provide a way for researchers to represent the quantity or category in scientific work—e.g., a categorical or continuous scale.
3. *Procedures*: describe what researchers need to do to carry out the measurement successfully.

Nevertheless, we must be clear that the way measuring is done can have implications well beyond the confines of the sciences, and for this reason scientific measures are likely to be hotly contested politically (Cartwright and Runhardt 2014). In the case of scientometrics, the scientists and their host institutions, e.g., universities and research institutes, are both aware that if their work is not classified as progressive, then the public and private organisations that fund them may well respond in specific ways that they don't want. In what follows, we describe some of the procedures that are available to manage potential bias in the measure of scientific progress from potential strategic reporting behaviour by researchers and their host institutions that can distort our measurements.

One of the main topics within scientometrics that has seen a huge investment of effort by ICT (information and computer technology) parties concerns the creation of citation indexes, released as commercial (e.g., Scopus, <https://www.scopus.com>) and even open services (e.g., OpenCitations, Peroni et al. 2015, <http://opencitations.net>). While counting citations is one of the most common and shared practices for assessing the quality of research—e.g., in several countries in Europe it has been used several times as one of the factors for assigning scientific qualifications to scholars—it is not the only one that can be considered for evaluating the quality of research. These additional assessment factors are usually classified according to two categories: (i) *intrinsic factors*, i.e., those related to the qualitative evaluation of the content of articles (quality of the arguments, identification of citation functions, etc.); and

(ii) *extrinsic factors*, i.e., those referring to quantitative characteristics of articles such as their metadata (number of authors, number of references, etc.) and other contextual characteristics (the impact of publishing venue, the number of citations received over time, etc.). Data Science technologies, including Machine Learning and Natural Language Processing tools, provide the grounds for automatizing the identification of these factors, such as the entities cited in articles (Fink et al. 2010), rhetorical structures (Liakata et al. 2010), arguments (Sateli and Witte 2015), and citation functions (Di Iorio et al. 2013).

The use of intrinsic-factor data can be very effective but also time consuming. They can be gathered manually by humans (e.g., through questionnaires to assess the intellectual perceptions of an article as in peer-review processes) as described in Opthof and colleagues (Opthof 2002). Other data of this specific kind can be extracted automatically by means of semantic technologies (e.g., machine learning, probabilistic models, deep machine readers) to retrieve, for instance, the functions of citations (i.e., author's reasons for citing a certain work) (Di Iorio et al. 2013).

Extrinsic factors, on the other hand, do not analyse the merit of a particular study considering its content; rather, they focus on using contextual data (such as citation counts) that should be able to predict, to some extent, the quality of the work in consideration. Thus, even if they are less accurate than the intrinsic factors, the extrinsic ones are usually preferred because they can be extracted in an automatic fashion by analysing papers, and they are available as soon as the paper is published. In addition to citation counts, other extrinsic factors can be: (i) the impact factor of the journals in which articles have been published, the number of references in articles, and the impact of the papers that have been cited by the articles in consideration, as introduced in Didegah and Thelwall (2013); (ii) the article length in terms of printed pages, as in Falagas et al. (2013); (iii) the number of co-authors and the rank of authors' affiliations according to QS World University Rankings, as in Antonakis et al. (2014); (iv) the number of bibliographic databases in which each journal of the selected articles was indexed, the proportion of the high-quality articles (measured according to specific factors) published by a journal and all the articles that have been published in the same venue in the same year independently from their quality, as in Lokker et al. (2008); (v) the price index—i.e., the percentage of papers cited by an article that have been published within 5 years before the publication year of such article, as in Onodera and Yoshikane (2014); and (vi) altimetrics about the papers—e.g., tweets, Facebook posts, Nature research highlights, mainstream media mentions and forum posts, as in Thelwall et al. (2013).

Donald Campbell (e.g., 1966) was one of the first to see the potential of unobtrusive measures in contexts where subjects were unlikely to offer unbiased responses to conventional data gathering procedures (e.g., questionnaires). However, perhaps only George Orwell (1949) could have imagined the breadth and depth of social science constructs that it is becoming possible to operationalise using data science tools.

Case 4: Strategy in the Tourism Sector

Tourism destinations are defined as complex amalgams of ‘products, amenities and services delivered by a range of highly interdependent tourism firms including transportation, accommodation, catering and entertainment companies and a wide range of public goods such as landscapes, scenery, sea, lakes, cultural heritage, socio-economic surroundings’ (Mariani 2016, p. 103). These elements are typically marketed and promoted holistically by local tourism organizations, conventions, and visitor bureaus. These are generally referred to as Destination Management Organizations (DMOs). More specifically, DMOs facilitate interactions and local partnerships between tourism firms for the development and delivery of a seamless experience that might maximize tourists’ satisfaction and the profitability of local enterprises. In continental Europe, most of the tourism destinations consist of Small and Medium Enterprises (SMEs) located in a specific geographical area that, on one hand, cooperate for destination marketing and product development purposes (under the aegis of a DMO) to increase inbound tourism flows and tourist expenditure while, on the other hand, they compete to win more customers (i.e., tourists and visitors) and profit from them.

This is the case of the Italian tourism sector where a high number of destinations consisting of a myriad of SMEs try to increase their market share of tourist arrivals, overnight stays, and tourism expenditure. Over the last three decades, globalization in travel and increased income allocated to travel have intensified competition between tourism destinations and among companies (Mariani and Baggio 2012; Mariani and Giorgio 2017). However, the most relevant driver of competitive advantage is technology development in ICTs (Mariani et al. 2014) that has brought about many different intermediaries (e.g., travel blogs, travelogues, online travel review sites, social media) for customers to share their opinions and reviews about destinations and tourism services in real time. The role played today by online travel review sites such as TripAdvisor or booking engines such as Booking and Expedia is becoming increasingly relevant as online ratings have been found to play a crucial role in pre-trip purchase decisions and to affect organizational performance measured through revenues and occupancy rates (Borghi and Mariani 2018).

Therefore, in addition to the traditional statistics related to arrivals, overnight stays in hotels, and accommodation facilities, DMOs today should deal with an increasing amount of unstructured data such as online searches, accommodation bookings, discussions, and images on social media produced by tourists and companies, as well as online consumer reviews.

However, DMOs as well as SMEs in the tourism and hospitality sector often ignore these data because they are unstructured, and therefore difficult to analyse and interpret. While individual SMEs have typically neither the budget nor the competences to deal with these data, only the most overfunded DMOs (in North America and Northern Europe) have equipped themselves with specific destination marketing systems that work in a similar way to enterprise resource planning systems. These platforms pool together data from both the supply (e.g., hotels, transportation companies, theme parks) and demand (e.g., bookings from prospective tourists) sides, matching them. Data science techniques are used to collect, analyse, process

(through online-analytical processing), report, and visualize data about the market trends, segments, evolution of bookings and occupancy rates, display offers of accommodation and transportation services as well as assemble accommodation, transportation, and other leisure activities (Mariani et al. 2018; Mariani and Borghi 2018).

However, it is still very difficult and complex to bring together the vast amount of structured and unstructured data produced before, during, and after visiting a destination. An interesting attempt has been carried out with the Destination Management Information System Åre (DMIS-Åre), developed by researchers of the Mid-Sweden University for the Swedish destination of Åre (Fuchs et al. 2014). The system consists of three sets of indicators: (i) economic performance indicators; (ii) customer behaviour indicators; and (iii) customer perception and experience indicators. The first group includes prices, bookings, reservations, hotel overnights, and so on. These data are relatively easy to extract. They are complemented with data about the users' behaviour: for instance, web navigation behaviours before reservations. It is particularly useful to have the analysis of booking channels and devices used for reservation. Customer behaviour indicators can be leveraged to identify clusters of tourists and create customized offers as well as identify and analyse trends, either historical or emergent. The last group of indicators includes information about the perception of the users and provides valuable indications about the destinations' attractiveness.

Building on the DMIS of Åre (Fuchs et al. 2014) and on an updated systematic review of the most relevant contribution at the intersection between Business Intelligence and *Big Data* in tourism and hospitality over the last 17 years (Mariani et al. 2018), we propose a prototype of a Destination Business Intelligence Unit (DBIU). The platform is useful for DMOs to: (i) improve the competitiveness of the destination (in terms of tourist arrivals and tourism expenditure as well as sustainability and carrying capacity); (ii) enhance the competitiveness of the SMEs operating in their hospitality sector. To this aim, our DBIU in addition to economic performance indicators, customer behaviour indicators, and customer perception and experience indicators adds sustainability and environmental indicators. Figure 1 summarizes our proposal and shows the relation with DMIS-Åre. The idea is to provide users with information about traffic and weather conditions, as well as consumption of electricity, gas, and water. These data can be used first to improve the users' experience by providing updated information in real time. In addition, data science techniques and tools can be used to better design and manage tourism services at the destination level by means of analysing tourists' preferences through their social media activity on smartphones and social location-based mobile marketing activities (Amaro et al. 2016; Chaabani et al. 2017).

Sustainability is increasingly important for today's destination managers and tourists, and can also be embedded in marketing and promotional strategies to attract green tourists (Mariani et al. 2016a, b) and improve the carrying capacity of the destination.

Moreover, our DBIU improves the "Functional, emotional value and satisfaction data" helping to enhance customer perception and experience indicators. The right

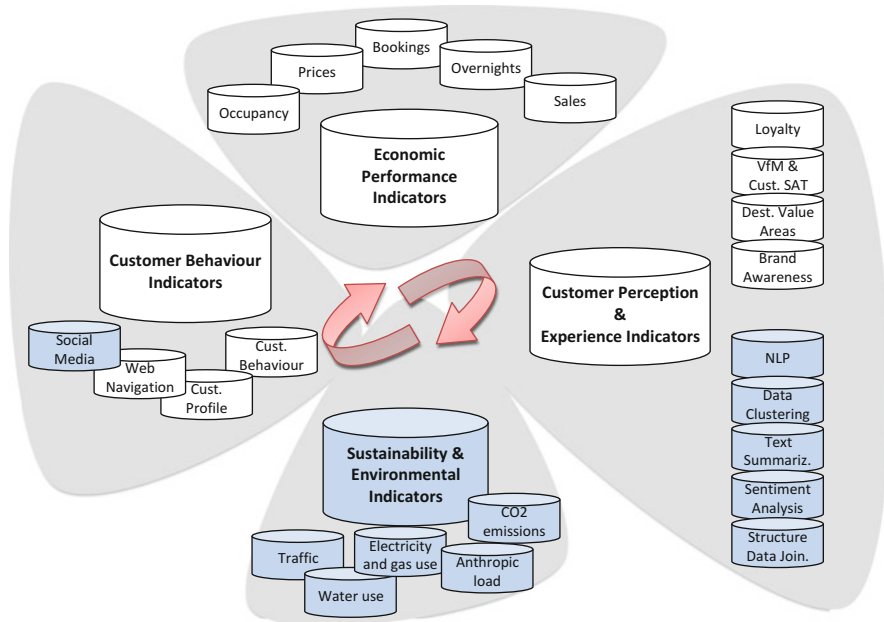


Fig. 1 Prototype of a DBIU

bottom part of the figure shows (in blue) our improvements. The primary goal is to analyse both structured and unstructured information by using modules of Natural Language Processing (NLP), text summarization, and sentiment analysis. The main data sources are the online reviews: they contain a significant amount of data but in different formats, languages and structures. Data science techniques can be exploited to (i) extract information from multiple sources (ii) define a common data model and normalize such heterogeneous information to that model, (iii) combine data into aggregated and parameterized forms, and (iv) visualize data in a clear way for the final customers. These techniques contribute to gaining a more comprehensive picture of users' perceptions.

As shown on the left-hand side of the picture, DBIU improves the customer behaviour indicators by leveraging a tool developed for data retrieval and analysis from the major social media. The tool consists of four modules, following the schema mentioned above: data extractor, parser, analyser, and visualizer modules (for a detailed description, see Mariani et al. 2016a, 2016b, 2017). That said, this DBIU might allow not only destination marketers and DMOs to match and process a vast amount of heterogeneous data but could also allow DMOs to share some of the relevant data related to customer behaviour and customer perceptions in real time with local SMEs operating in the accommodation and transportation industries. While this prototype could certainly be the object of further improvement, we believe that it represents an interesting tool to strengthen local economic systems heavily reliant on tourism.

Case 5: Business Performance Analytics

Current competitive marketplaces are “hyper-challenging” for organizations in a continued search for opportunities to maintain and improve business growth and profitability. In this context, management control systems play an important role to support management by providing key information and quick feedback for strategic and operational decision-making.

Technology is changing the rules of business and how to transform data into knowledge has become a key issue (Davenport et al. 2010). There is a growing consensus that business analytics and *Big Data* have huge potential for performance management (Bhimani and Willcocks 2014), informing decision-making, and improving business strategy formulation and implementation (CIMA 2014). Such potential has been generally acknowledged by the literature; however, organizations report significant difficulties in extracting strategically valuable insights from data (CIMA 2014).

Progress in ICT has opened new opportunities in terms of modelling organisational operations and managing firms in real time and has attracted interest in the relations between control and information systems (Dechow et al. 2007). While information systems have been considered important enablers of performance management, their role is not yet understood either theoretically or practically (Nudurupati et al. 2011, 2016). Indeed, several questions arise. A key issue concerns the analysis of data availability and sources (Zhang et al. 2015). Secondly, quantity and variety bring additional concerns in terms of data quality and relevance (IFAC 2011; Bhimani and Willcocks 2014). As for the former, organizations have access to an unprecedented amount of data and to previously unimaginable opportunities to analyse them. ICT represents a strategic success factor because of its potential to collect and offer such huge amounts of data. As for the latter, while the availability of data does not necessarily mean information, the ability to understand and extract value from them becomes critical too. From this perspective, Business Performance Analytics (BPAs) offer valuable support (Silvi et al. 2012) because they link data collection and use to a previous understanding of an organization’s business model, its deployment into key success factors and performance measures, and finally performance management routines.

Consistent with the literature, this fifth case focuses on the challenging relationship between BPA and ICT and its effect on their abilities to collect, select, manage, and interpret data. Specifically, it highlights the key issues that arise when integrating the use of BPA within the performance measurement and management process, in the light of the support provided by ICT in (i) automatic data collection (i.e., tools able to extract a large amount of data from multiple heterogeneous sources), (ii) data analysis (i.e., tools combining machine-learning data warehouse and (iii) decision-making techniques to identify patterns and trends) and data visualization (i.e., novel interfaces and paradigms make data available and easier to consume).

BPA refers to the extensive use of multiple data sources and analytical methods to drive decisions and actions, by understanding and controlling business dynamics and performance (Davenport and Harris 2007, p. 7) and supporting effective PMS design and adoption (Silvi et al. 2012). Examples are decision support systems, expert

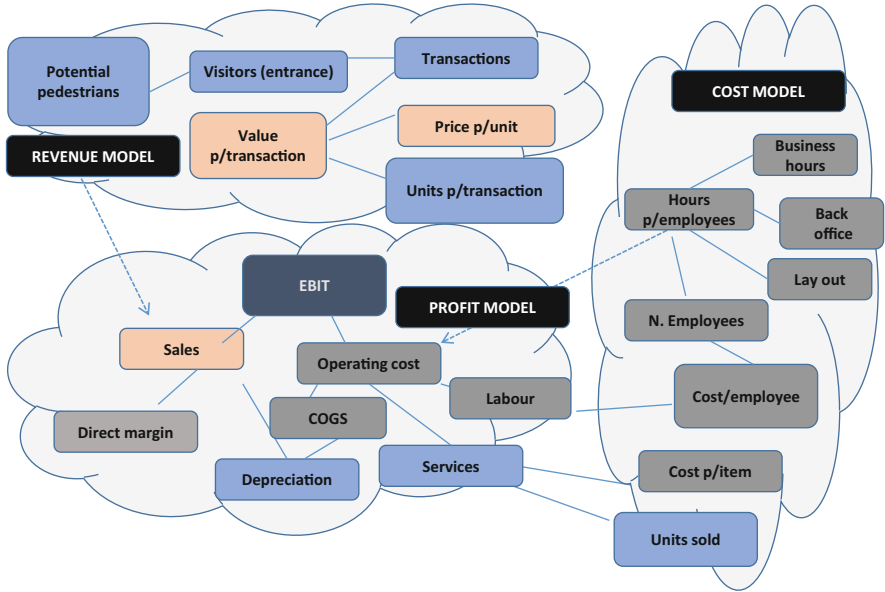


Fig. 2 Bookstore performance map

systems, data mining systems, probability modelling, structural empirical models, optimization methods, explanatory and predictive models, and fact-based management. BPA are then focused around management needs and their design requires (i) the comprehension of a company’s business model and context, and the way its performance is achieved, (ii) the identification of key success factors, information needs, data sources, (iii) the provision of an information platform and analytical tools (descriptive, exploratory, predictive, prescriptive, cognitive); (iv) the assessment of performance factors and drivers, and (v) the visualization of business performance and dynamics and their management.

Figure 2 shows an example of a business performance map of a bookstore. Specifically, business profitability (EBIT) is the result of the company’s revenues and cost model. Revenues—driven at a first level by price and unit sold—can be further broken down, showing the most elementary revenue drivers: people flow, entrance rate, and conversion rate, and purchase. On the other hand, costs are driven by volumes, product categories, and related cost, as well as by activity hours (labour), shop layout (efficiency), sourcing factors (delivery time), etc. Gauging these dynamics and their factors allows the store manager to understand better the way performance is achieved and can be improved.

On the other hand, this performance and measurement system requires data availability, data analysis, data visualization technologies, analytical methods, routines and performance management skills, and attitudes and talents. Hence, the implementation of BPA and analytical Business Performance Management systems

is by nature a complex task, as it involves managerial, analytical, and ICT competencies and tools. From a technological point of view, there are at least three main challenging steps: (i) data collection, (ii) data analysis, and (iii) data visualization.

Data collection. Data originates from different internal and external sources, and are stored in several systems, with different languages and forms (conversational, video, text, etc.), timing, size, accuracy, and usability (open- and closed-access). Particularly interesting is the integration of structured data with unstructured and semi-structured data, extracted from documents, which represent a huge source of knowledge and competitive assets made available by Natural Language Processing techniques (Cambria and White 2014). As discussed by Zhang et al. (2015), some specific features of digital and *Big Data* challenge the capabilities of modern information systems; they are known as the 4 Vs: huge Volume, high Velocity, huge Variety, and uncertain Veracity. Despite the mentioned potential benefits, then, these critical issues still undermine ITC systems' effectiveness for BPA purposes (Beaubien 2012) and a number of questions arise. How to collect data? How to blend them? What about data security?

Data Analysis. This concerns the choice of the analytical method (descriptive, exploratory, predictive, prescriptive, and cognitive). From a technical point of view, key issues are how to use data for those typologies of analytics and how to design expressive data models. The interaction between domain experts and technical experts is crucial to achieve this goal. Another key issue is the integration between different models (for instance, predictive, prescriptive or cognitive models) and techniques to combine data, such as embedded analytics, machine learning, artificial intelligence, data warehousing, and data mining (Han et al. 2011; Kimball and Ross 2011). Automatic reasoning and decision-making about data complete the path.

Data visualization. The challenge is how to report the analytical and performance infrastructure into visual formats easy to access and understand, aligned with user experience and expectations. The success factor is not only to aggregate data but also to extract unexpected and hidden information and trends.

To summarize, in an age of digital economy, a successful contribution of performance management systems and ICT to business competitiveness and innovation is undoubtedly interrelated and their effective implementation requires a holistic approach. Achieving competitive advantage with analytics requires a change in the role of data in decision-making that involves information management and cultural norms (Ransbotham et al. 2016). Another issue is about analytics talent, in form of "translators," as first, able to bridge IT and data issues to decision making with a contribution to the design and execution of the overall data-analytics strategy while linking IT, analytics, and business-unit teams. Furthermore, data scientists should combine strong analytics skills with IT know-how, driving towards sophisticated models and algorithms. Because digital skills and talents are scarce, they represent an opportunity for research and education, and value for community wellbeing.

3 Conclusions and Implications for Entrepreneurship Research

Entrepreneurship research covers a rather wide range of problems, contexts and processes, usually combining different social science perspectives. Generating better understanding and theories about entrepreneurial and value-creation processes is challenging because new ventures develop different internal resources and characteristics, evolve under changing external environments and pursue business ideas that are changing over time. Collecting primary data for quantitative studies covering all these aspects is extremely time consuming and resource demanding. It involves mapping of the individual entrepreneurs, their ventures, and the external environment from inception and over a significant period of time until the venture has reached a mature stage. Some of these challenges can be overcome by tapping into the increasingly rich sources of digital data that are being generated about the activities of individuals, firms, and their contexts. Making use of data science and ICT tools is necessary to tap into and refine these data sources.

While the technical availability of databases and their subsequent commercial development in the seventies and eighties opened numerous opportunities to access longitudinal and structured data, their level of specification and detail has been inadequate on many grounds (too general, incomplete, self-reported, etc.). Data gathering, storage, and manipulation have therefore become a key element in any research program, but often with inefficient replication of efforts and low levels of sharing to allow for proper replicability or further enhancement of analyses. The evolution in data science technologies and research opportunities are becoming pervasive in many different types of research and approaches as illustrated by the various cases presented in this chapter. We are at the break of a new dawn for reconsidering the use of field data in entrepreneurship research.

First, its ubiquitous nature calls for creativity in designing new approaches to collect evidence as traces in a field track, left there not to mark the trail, but simply because of walking. And yet, as much as zoologist and anthropologists have used tracks to understand migration patterns and their evolutionary consequences, several digital marks can have a profound relevance to understand individual and collective behaviour and their implications for entrepreneurship. Case 1 offered us a specific example associated with the analysis of interpersonal networks that clearly has the potential to inspire new ways of collecting data on the characteristics and performance of entrepreneurs and entrepreneurial teams.

Second, the possibility of standardizing the data-gathering procedure in multiple geographical locations could help overcome significantly the current limitations of pursuing comparative analysis in different countries and settings. Although interoperability standards and data coding procedures are still far from allowing for a frictionless aggregation, the progress in these areas has shown clear opportunities, as Case 4 exemplified in the field of tourism. Many industries increasingly rely on digital platforms for key business processes, which provides new opportunities to shed light on the role of entrepreneurship in these industries.

Third, new and original datasets could come from the aggregation of existing sources and be designed from the beginning as able to automatically or semi-automatically update to continue providing users with both historical accounts and the most recent evidence. Cases 2 and 3 discussed examples related to datasets of different sizes, compositions, and spans, ranging from research driven to institutionally and commercially driven ones. Such combinations of data sources to trace entrepreneurial efforts over time will be highly valuable for generating a better understanding of the entrepreneurial process and the resulting outcomes and impacts of entrepreneurship at different levels of analysis.

Fourth, and probably more evident in its short-term impact, decision making processes, tools and roles are being revolutionized in many organizations and will soon impact all of us in direct or indirect ways. Business analysis and intelligence, as described by Case 5, are two areas where the attention of entrepreneurship scholars have long focused to identify the sources of competitive advantage, map the evolution of organizational complexity over the life of new ventures, or assess the differences (if any) between managers and entrepreneurs. Clearly, digitalization and the use of data science not only provide new opportunities for academics, but are also profoundly influencing the opportunities of entrepreneurship in many areas of society. Hence, entrepreneurship scholars experience changes in both the empirical phenomenon as well as the data and methods available, driven by data science.

This chapter has been written as a collaborative effort between data scientists and management scholars and thereby illustrates the need for cross- and multi-disciplinary approaches to fully benefit from the rapidly increasing access to big data. And yet, the more we try to link what has been presented by many creative scholars in this chapter as new ideas to productively and creatively match entrepreneurship and data science research, the more additional ideas are emerging. We are looking forward to reading ideas from other scholars and we hope this chapter has offered some inspirations to begin an exciting and unpredictable new journey.

Acknowledgment This chapter is an outcome of the research workshop entitled *Management meet ICT and return: New trends in data science*, which was held on 25th October 2016 at the Department of Management of the University of Bologna in Italy with the support of the FP7-PEOPLE-2011-CIG-303502-TASTE Project (www.project-taste.eu).

References

- Amaro, S., Duarte, P., & Henriques, C. (2016). Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, 59, 1–15.
- Anderson, A., Dodd, S. D., & Jack, S. (2010). Network practices and entrepreneurial growth. *Scandinavian Journal of Management*, 26(2), 121–133.
- Ankrah, S., & AL-Tabbaa, O. (2015). Universities-industry collaboration: A literature review. *Scandinavian Journal of Management*, 31(3), 387–408.
- Antonakis, J., Bastardo, N., Liu, Y., & Schriesheim, C. A. (2014). What makes articles highly cited? *The Leadership Quarterly*, 25(1), 152–179. <https://doi.org/10.1016/j.leaqua.2013.10.014>.

- Arrow, T., & Kasberger, S. (2017). *Introduction to content mine: Tools for mining scholarly and research literature*. Virginia Tech. University Libraries. <http://hdl.handle.net/10919/77525>
- Beaubien, L. (2012). Technology, change, and management control: A temporal perspective. *Accounting, Auditing & Accountability Journal*, 26(1), 48–74.
- Bhimani, A., & Willcocks, L. (2014). Digitisation, “Big Data” and the transformation of accounting information. *Accounting and Business Research*, 44(4), 469–490.
- Borghì, M. & Mariani, M. (2018). Electronic word of mouth and firm performance: Evidence from the hospitality sector. *Tourism Management*, forthcoming.
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, 9(2), 48–57.
- Cartwright, N. (2014). Causal inference. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction*. Oxford: Oxford University Press.
- Cartwright, N., & Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction*. Oxford: Oxford University Press.
- Chaabani, Y., Toujani, R., & Akaichi, J. (2017, June). Sentiment analysis method for tracking tourists reviews in social media network. In *International conference on intelligent interactive multimedia systems and services* (pp. 299–310). Cham: Springer.
- CIMA. (2014). *Big Data. Readyng business for the Big Data revolution*. [online] Accessed July 15, 2015, from <http://www.cgma.org/Resources/Reports/DownloadableDocuments/CGMA-briefing-big-data.pdf>
- Cygniak, R., Wood, D., & Lanthaler, M. (2014). *RDF 1.1 Concepts and abstract syntax*. W3C Recommendation 25 February 2014. Accessed July 20, 2017, from <http://www.w3.org/TR/rdf11-concepts/>
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Boston: Harvard Business Press.
- Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions: Better results*. Boston: Harvard Business Press.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65(3), 122–135.
- Dechow, N., Granlund, M., & Mouritsen, J. (2007). Interactions between information technology and management control. In D. Northcott et al. (Eds.), *Issues in management accounting* (3rd ed.). London: Pearson.
- Di Iorio, A., Nuzzolese, A. G., & Peroni, S. (2013). Characterising citations in scholarly documents: The CiTalO framework. *ESWC (Satellite Events), 2013*, 66–77. https://doi.org/10.1007/978-3-642-41242-4_6.
- Didegah, F., & Thelwall, M. (2013). Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, 64(5), 1055–1064. <https://doi.org/10.1002/asi.22806>.
- Droll, A., Shahzad, K., Ehsanullah, E., & Stoyan, T. (2017). Using artificial intelligence and web media data to evaluate the growth potential of companies in emerging industry sectors. *Technology Innovation Management Review*, 7(6), 25–37.
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research Policy*, 29(2), 109–123.
- Falagas, M. E., Zarkali, A., Karageorgopoulos, D. E., Bardakas, V., & Mavros, M. N. (2013). The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals. *PLoS One*, 8(2), e49476. <https://doi.org/10.1371/journal.pone.0049476>.
- Ferrucci, D., Lally, A., Verspoor, K., & Nyberg, E. (2009). *Unstructured information management architecture (UIMA) version 1.0. OASIS standard*. Accessed July 29, 2017, from <http://docs.oasis-open.org/uima/v1.0/uima-v1.0.html>
- Fini, R., & Toschi, L. (2016). Academic logic and corporate entrepreneurial intentions: A study of the interaction between cognitive and institutional factors in new firms. *International Small Business Journal*, 34(5), 637–659.

- Fini, R., Fu, K., Mathisen, M. T., Rasmussen, E., & Wright, M. (2017). Institutional determinants of university spin-off quantity and quality: A longitudinal, multilevel, cross-country study. *Small Business Economics*, 48(2), 361–391.
- Fink, J. L., Fericola, P., Chandran, R., Parastatidis, S., Wade, A., Naim, O., Quinn, G. B., & Bourne, P. E. (2010). Word add-in for ontology recognition: Semantic enrichment of scientific literature. *BMC Bioinformatics*, 2010(11), 103. <https://doi.org/10.1186/1471-2105-11-103>.
- Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big Data analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management*, 3(4), 198–209.
- George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and data science methods for management research. *Academy of Management Journal*, 59(5), 1493–1507.
- Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Burlington: Elsevier.
- Hardwick, J., Anderson, A. R., & Cruickshank, D. (2013). Trust formation processes in innovative collaborations: Networking as knowledge building practices. *European Journal of Innovation Management*, 16(1), 4–21.
- Harris, S., & Seaborne, A. (2013). *SPARQL 1.1 Query Language*. W3C Recommendation 21 March 2013. Accessed July 28, 2017, from <http://www.w3.org/TR/sparql11-query/>
- IFAC (2011). *Predictive business analytics: Improving business performance with forward-looking measures*. [online] Accessed July 26, 2017, from <https://www.ifac.org/publications-resources/predictive-business-analytics-improving-business-performance-forward-looking>
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: The complete guide to dimensional modeling*. New York: Wiley.
- Liakata, M., Teufel, S., Siddharthan, A., and Batchelor, C. (2010). In Proceedings of the 7th international conference on language resources and evaluation (LREC 2010): 2054–2061. http://www.lrec-conf.org/proceedings/lrec2010/pdf/644_Paper.pdf
- Lokker, C., McKibbin, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: Retrospective cohort study. *BMJ*, 336(7645), 655–657. <https://doi.org/10.1136/bmj.39482.526713.BE>.
- March, J. G., & Simon, H. A. (1958). *Organizations*. Oxford: Wiley.
- Mariani, M. M. (2016). Coordination in inter-network co-opetition: Evidence from the tourism sector. *Industrial Marketing Management*, 53, 103–123. <https://doi.org/10.1016/j.indmarman.2015.11.015>.
- Mariani, M. M., & Baggio, R. (2012). Special Issue: Managing Tourism in a Changing World: Issues and Cases. *Anatolia: An International Journal of Tourism and Hospitality Research*, 23(1), 1–3. <https://doi.org/10.1080/13032917.2011.653636>.
- Mariani, M. M., & Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management*, 66, 47–52. <https://doi.org/10.1016/j.tourman.2017.11.006>.
- Mariani, M. M., & Giorgio, L. (2017). The “Pink Night” festival revisited: Meta-events and the role of destination partnerships in staging event tourism. *Annals of Tourism Research*, 62(1), 89–109.
- Mariani, M. M., Baggio, R., Buhalis, D., & Longhi, C. (2014). *Tourism management, marketing, and development: The importance of networks and ICTs* (Vol. I). New York: Palgrave Macmillan. <https://doi.org/10.1057/9781137354358>.
- Mariani, M. M., Di Felice, M., & Mura, M. (2016a). Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, 54, 321–343. <https://doi.org/10.1016/j.tourman.2015.12.008>.
- Mariani, M. M., Czakon, W., Buhalis, D., & Vitouladiti, O. (2016b). Introduction. In M. M. Mariani, W. Czakon, D. Buhalis, & O. Vitouladiti (Eds.), *Tourism management, marketing,*

- and development* (pp. 1–12). Palgrave Macmillan: New York. https://doi.org/10.1057/9781137401854_1.
- Mariani, M. M., Mura, M. & Di Felice, M. (2017). The determinants of Facebook social engagement for national tourism organisations' Facebook pages: A quantitative approach. *Journal of Destination Marketing & Management*, forthcoming. <https://doi.org/10.1016/j.jdmm.2017.06.003>.
- Mariani, M. M., Baggio, R., Fuchs, M. & Höpken, W. (2018). Business intelligence and big data in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management*, forthcoming.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90, 61–67.
- McPhee, C. (2016). Editorial: Managing innovation. *Technology Innovation Management Review*, 6(4), 3–4.
- Morse, E. A., Fowler, S. W., & Lawrence, T. B. (2007). The impact of virtual embeddedness on new venture survival: Overcoming the liabilities of newness. *Entrepreneurship: Theory and Practice*, 31(2), 139–159.
- Motik, B., Patel-Schneider P. F., and Parsia B. (2012). *OWL 2 web ontology language - Structural specification and functional-style syntax* (2nd ed.). W3C recommendation 11 December 2012. Accessed July 28, 2017, from <http://www.w3.org/TR/owl2-syntax/>
- Nayyar, P. R. (1990). Information asymmetries: A source of competitive advantage for diversified service firms. *Strategic Management Journal*, 11(7), 513–519.
- Nudurupati, S. S., Tebboune, S., & Hardman, J. (2016). Contemporary performance measurement and management (PMM) in digital economies. *Production Planning & Control*, 27(3), 226–235.
- Nudurupati, S. S., Bititci, U. S., Kumar, V., & Chan, F. T. S. (2011). State of the art literature review on performance measurement. *Computers and Industrial Engineering*, 60, 279–290.
- Onodera, N., & Yoshikane, F. (2014). Factors affecting citation rates of research articles: Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, 66(4), 739–764. <https://doi.org/10.1002/asi.23209>.
- Ophthof, T. (2002). The significance of the peer review process against the background of bias: Priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovascular Research*, 56(3), 339–346. [https://doi.org/10.1016/S0008-6363\(02\)00712-5](https://doi.org/10.1016/S0008-6363(02)00712-5).
- Orwell, G. (1949). *1984*. New York: Houghton Mifflin Harcourt.
- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy--and how to make them work for you*. New York: WW Norton & Company.
- Perkmann, M., Neely, A., & Walsh, K. (2011). How should firms evaluate success in university-industry alliances? A performance measurement system. *R&D Management*, 41(2), 202–216.
- Perkmann, M., Fini, R., Ross, J. M., Salter, A., Silvestri, C., & Tartari, V. (2015). Accounting for universities' impact: Using augmented data to measure academic engagement and commercialization by academic scientists. *Research Evaluation*, 24(4), 380–391.
- Peroni, S., Dutton, A., Gray, T., & Shotton, D. (2015). Setting our bibliographic references free: Towards open citation data. *Journal of Documentation*, 71(2), 253–277.
- Polanyi, M. (2013). *The tacit dimension*. Garden City, NY: Doubleday.
- Powell, W., Koput, K., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1), 116–145.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to Big Data and data-driven decision making. *Big Data*, 1(1), 51–59.
- Ransbotham, S., Kiron, D. & Kirk Prentice, P. (2016). Beyond the hype: The hard work behind analytics success. *The 2016 Data & Analytics Report by MIT Sloan Management Review & SAS*.

- [online] Accessed July 14, 2016, from <http://sloanreview.mit.edu/projects/the-hard-work-behind-data-analytics-strategy/>
- Ritter, T., & Gemunden, G. (2003). Interorganizational relationships and networks: An overview. *Journal of Business Research*, 56(9), 691–697.
- Sateli, B., & Witte, R. (2015). Semantic representation of scientific literature: Bringing claims, contributions and named entities onto the Linked Open Data cloud. *PeerJ Computer Science*, e37. <https://doi.org/10.7717/peerj-cs.37>.
- Shotton, D. (2009). Semantic publishing: The coming revolution in scientific journal publishing. *Learned Publishing*, 22(2), 85–94.
- Sigfusson, T., & Chetty, S. (2013). Building international entrepreneurial virtual networks in cyberspace. *Journal of World Business*, 48(2), 260–270.
- Silvi, R., Bartolini, M., Raffoni, A., & Visani, F. (2012). Business performance analytics: Level of adoption and support provided to performance measurement systems. *Management Control*, 3 (Special Issue), 117–142.
- Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, 1(1), 161–176.
- Sundararajan, A. (2016). *The sharing economy: The end of employment and the rise of crowd-based capitalism*. Cambridge: MIT Press.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS One*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>.
- Tidd, J. (2001). Innovation management in context: Environment, organization and performance. *International Journal of Management Reviews*, 3(3), 169–183.
- Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communication of the ACM*, 57(10), 78–85.
- West, J., Vanhaverbeke, W., & Chesbrough, H. (2006). Open innovation: A research agenda. In H. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), *Open innovation: Researching a new paradigm* (pp. 285–307). Oxford: Oxford University Press.
- Williams, A. J., Harland, L., Groth, P., Pettifer, S., Chicheste, C., et al. (2012). Open PHACTS: Semantic interoperability for drug discovery. *Drug Discovery Today*, 17(21-22), 1188-1198.
- Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *The journal of Law and Economics*, 22(2), 233–261.
- Zhang, C. (2015). *DeepDive: A data management system for automatic knowledge base construction*. Ph.D. Dissertation, University of Wisconsin-Madison. Accessed July 29, 2017, from <http://cs.stanford.edu/people/czhang/zhang.thesis.pdf>
- Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward effective big data analysis in continuous auditing. *Accounting Horizons*, 29(2), 469–476.
- Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class Hadoop and streaming data*. New York: McGraw-Hill.