Book
Rethinking entrepreneurship education:
the role of collaborative practices and innovation

Editor
Springer

Book series
Entrepreneurship, Structural change and Industrial Dynamics

Guest editors
G. Bosio, T. Minola, F. Origo and S. Tomelleri

Publication date
2018

Chapter #4
Collaborative practices and multidisciplinary research: the dialogue between entrepreneurship, management and data science

Contributors
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*

* Department of Management, University of Bologna, Italy
** Imperial College Business School London, UK
*** Department of Computer Science and Engineering, University of Bologna, Italy
**** Nord University Business School Bodø, Norway
***** Ratio Institute, Stockholm, Sweden
******Henley Business School, University of Reading, UK

Corresponding author
riccardo.fini@unibo.it

Version
02/12/2017

Acknowledgment
This chapter is an outcome of the research workshop entitled *Management meet ICT and return: new trends in data science,* 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).

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 a variety of fields are coming into the spotlight not only from scholarly communities but also the popular press. To join this conversation, 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 evidence arising from different collaborative projects addressing: university-industry collaborations, the impact of technology-based activities, the measurement of scientific productivity, as well as performance measurement and business analytics. Implications for collaborative practices in entrepreneurship research are discussed.

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 science 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 customer 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 is generated continuously, both inside and outside the Internet. Every digital process and economic transaction produces some 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[1] and natural language processing[2] (NLP), as well as

[1] Machine learning: Technology 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.
[2] Natural Language Processing: Technology 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.

on cloud-based applications[3]. Computers running machine learning or NLP algorithms can help to 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 endeavors, new opportunities are also opening for unexplored alternatives in other research domains where classification, parsing, and clustering of text and images has so far dependent mostly on human centered activities and interpretation. Management and entrepreneurial 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, Van Alstyne and Choudary (2016) analyze how two-sided network effects can be leveraged to build effective cloud-based product platforms, showing once more how data-driven technologies can be key determinants of competitive advantage. A similar conclusion is reached by Arun Sundararajan (2016) 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. George et al. (2016) in a recent editorial published in the Academy of Management Journal discuss at length how to frame the challenges faced. More precisely, they suggest distinguishing between the affects 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 analyze 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 since the early nineties significantly changed the field. However, these new opportunities are still far from being incorporated into doctoral programs for the new generations of researchers to become familiar with them and certainly require the education of many editors to be able to properly staff their reviewing teams to make sure 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 several 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,

[3] Cloud-based applications: Companies can take advantage of the elastic nature of the cloud and deploying their products 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 same environment, multiple enterprise environments remain in sync, and their flows of data can be easily integrated. Since applications in the cloud are always deployable, always available, and highly scalable, continuous, agile innovation becomes an objective achievable by any business.

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 it takes to gather and analyze the necessary information. Opportunistic behaviors 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,[4] 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. To advance our understanding of these new possibilities, in the next section we explore some preliminary ideas originated within five different collaborative projects, operating at the interface between management and data science research. 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, behaviors, and actions in a manner that was not previously possible. Second, we look at the case of technology innovation management, for instance, to measure the impact of technology-based activities, e.g., the patent protection of intellectual property rights, that 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-)automatic creation of structured data from non-structured content and generate meaningful interlinks. Third, 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 need to 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, this case combines 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, 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 amount of destinations as well as SMEs often ignore or underuse this type of data because it is unstructured and therefore difficult to analyze and interpret. Several applications, developed to solve different problems, could offer viable opportunities to overcome these limitations and strengthen local economic systems. Fifth, our and last case takes the collaborations between management and ICT one step further. Specifically, it explores the role of business performance

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

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.

Cross-fertilizations between management 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 an increasing number of studies investigate a variety of questions in the field.

Ankraa and Tabbaa (2016) propose a conceptual framework to analyse UIC by identifying five key aspects that emerged from a systematic review of the literature, highlighting 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 (Perkmann et al., 2013) 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.

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, Dodd and Jack 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 SNSs links is 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 (1967) 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 the policy makers. 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 the development of 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, offers 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 of 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, analyzing firms adopting a business model that encourages open collaboration;

and economic development in a knowledge-based society (McPhee, 2016) are some typically investigated topics in this academic 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 behaviors, organization 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 startups 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 is 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, and they are supported by quite 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 fifteen years, web technologies have been radically expanded and now they include several languages and data models that allow anyone to make available structured data on the most disruptive communication platform in recent decades, i.e. the web. These new tools, named as *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 anymore in monolithic silos, which usually happens with common databases, but 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, the 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 been started for making available scholarly-related data on the web by means of semantic web formats, such as OpenCitations (http://opencitations.net), that publishes citation data (Peroni et al., 2015), Open PHACTS (https://www.openphacts.org/), that makes available data about drugs (Williams et al., 2012), and Wikidata (https://wikidata.org), which contains encyclopedic 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, which 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 the Natural Language Processing (NLP), which is the process of 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 analyze the *innovative productivity* of Italian public research institutes. Rather than focusing on single patent documents, PATIRIS allows users to analyze 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 an 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 5 different domains at the individual-, knowledge-, firm-, institutional- and contextual-level. More precisely, it analyzes about 60,000 academics, 1,000 patents, 1,100 spin-offs and 95 universities in 20 regions for the period 2000-2014. To obtain this multilevel structure the researchers integrated data derived from ad-hoc surveys to research offices, technology transfer offices, spin-offs and entrepreneurs, the Cineca dataset and LinkedIn for individual data, the European Patent Office and PATIRIS for patent data, the MIUR dataset for institutional data, Eurostat for contextual data at the regional level and others. In this research design, carefully automated and structured retrieval, import, cleaning and integration of the data are clearly critical for the integrity of the data and the feasibility of the project.

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 by means of 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 on 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 5, as opposed to 4 on a clearly articulated anchoring scale. The accumulating of publications and citations, corrected for self-citation and weighted for author affiliation, 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 predict 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 & Runhardt, 2014):

1) *Characterisation*, layout 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 be hotly contested politically (Cartwright & 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 it in specific ways that they don't want. In what follow 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 the scientific habitation to scholars – it is not the only one that can be considered for evaluating the quality of research works. These additional assessment factors are usually classified according to two categories: i) *intrinsic factors*, i.e., those related with the qualitative evaluation of the content of articles (quality of the arguments, identification of citation functions, etc.); 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 citation received during 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 factors 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 in Opthof and colleagues (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), so as to retrieve, for instance, the functions of citations (i.e., author's reasons for citing a certain) (Di Iorio et al., 2013).
Extrinsic factors, on the other hand, do not analyse the merit of a particular research work considering its content, but rather they focus in 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 one are usually preferred since they can be extracted in an automatic fashion by analysing papers, and that they are available as soon as the paper is published in some venue. In addition to citation counts, other extrinsic factors can be: i) the impact factor of the journals where 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 and colleagues (2013); iii) the number of co-authors and the rank of authors' affiliations according to QS World University Rankings, as in Antonakis and colleagues (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 factor) 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 and colleagues (2008); v) the price index, i.e. the percentage of the papers cited by an article that have been published within five years before the publication year of such article, as in Onodera and Yoshikane (2014); vi) altimetric about the papers, e.g. tweets, Facebook posts, Nature research highlights, mainstream media mentions and forum posts, as in Thelwall and colleagues (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

Tourism destinations are defined in tourism management literature 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, 2015: p. 103). These elements are typically marketed and promoted holistically by local tourism organizations and conventions, as well as visitor bureaus. These are generally referred 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 with each other 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 in terms 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 such 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 (Xiang et al., 2017) and to affect organizational performance measured through revenues and occupancy rates (Viglia et al, 2016)

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 SME in the tourism and hospitality sector often ignore this type of data, precisely because it is unstructured, and therefore difficult to analyze 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 the supply side (e.g., hotels, transportation companies, theme parks) and from the demand side (e.g., bookings from perspective tourists) and match them. Data science techniques are used to collect, analyze, 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., 2017).

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 colleagues 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 behavior indicators; 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' behavior, for instance web navigation behaviors before reservations. It is particularly useful to have the analysis of booking channels and devices used for reservation. Customer behavior indicators can be leveraged to identify clusters of tourists and create customized offers as well as identify and analyze 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. The overall framework is shown in the following picture, taken from the original paper.

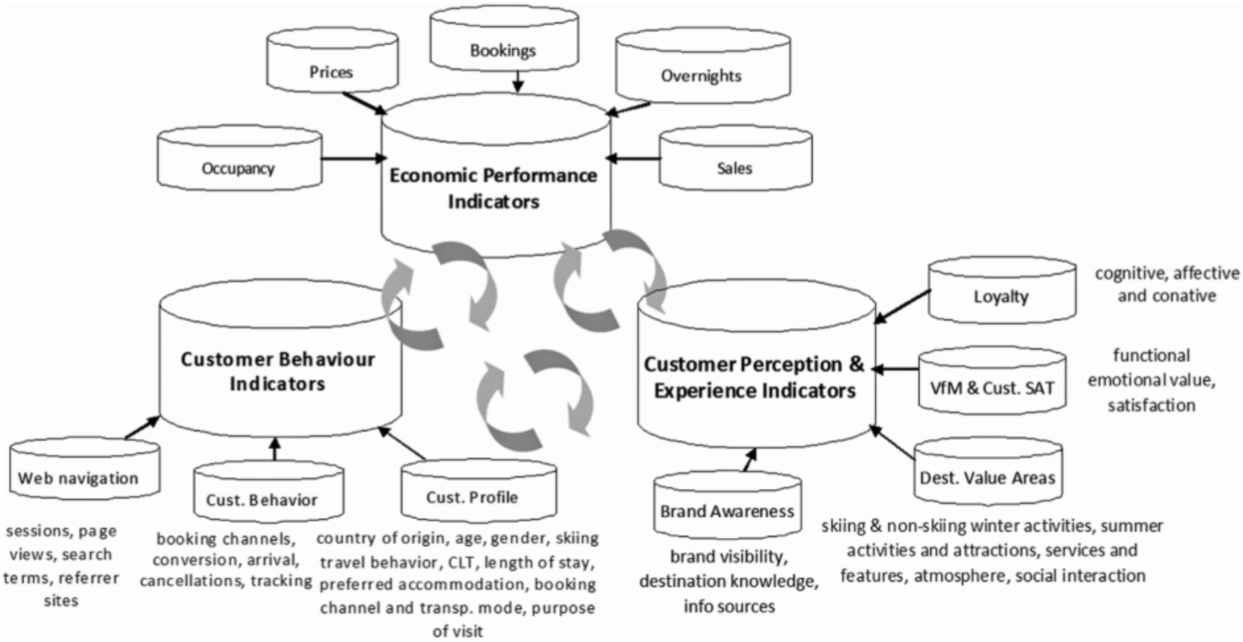

Fig. 1 DMIS-Åre (Fuchs et al., 2014)

Building on the DMIS (Fuchs, 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., 2017), we propose a prototype of a Destination Business Intelligence Unit (DBIU). The platform is useful for DMOs to: 1) improve the competitiveness of the destination (in terms of tourist arrivals and tourism expenditure as well as sustainability and carrying capacity); 2) enhance the competitiveness of the SMEs operating in their hospitality sector. To this aim our DBIU in addition to economic performance indicators, customer behavior indicators and customer perception & experience indicators, adds sustainability and environmental indicators. Figure 2 summarizes our proposal and shows the relation with DMIS-Åre.

DBIU in addition to economic performance indicators, customer behavior indicators and customer perception & experience indicators, adds sustainability and environmental indicators. 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 analyzing 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 (Ogonowska and Torre, 2016) and improve the carrying capacity of the destination.

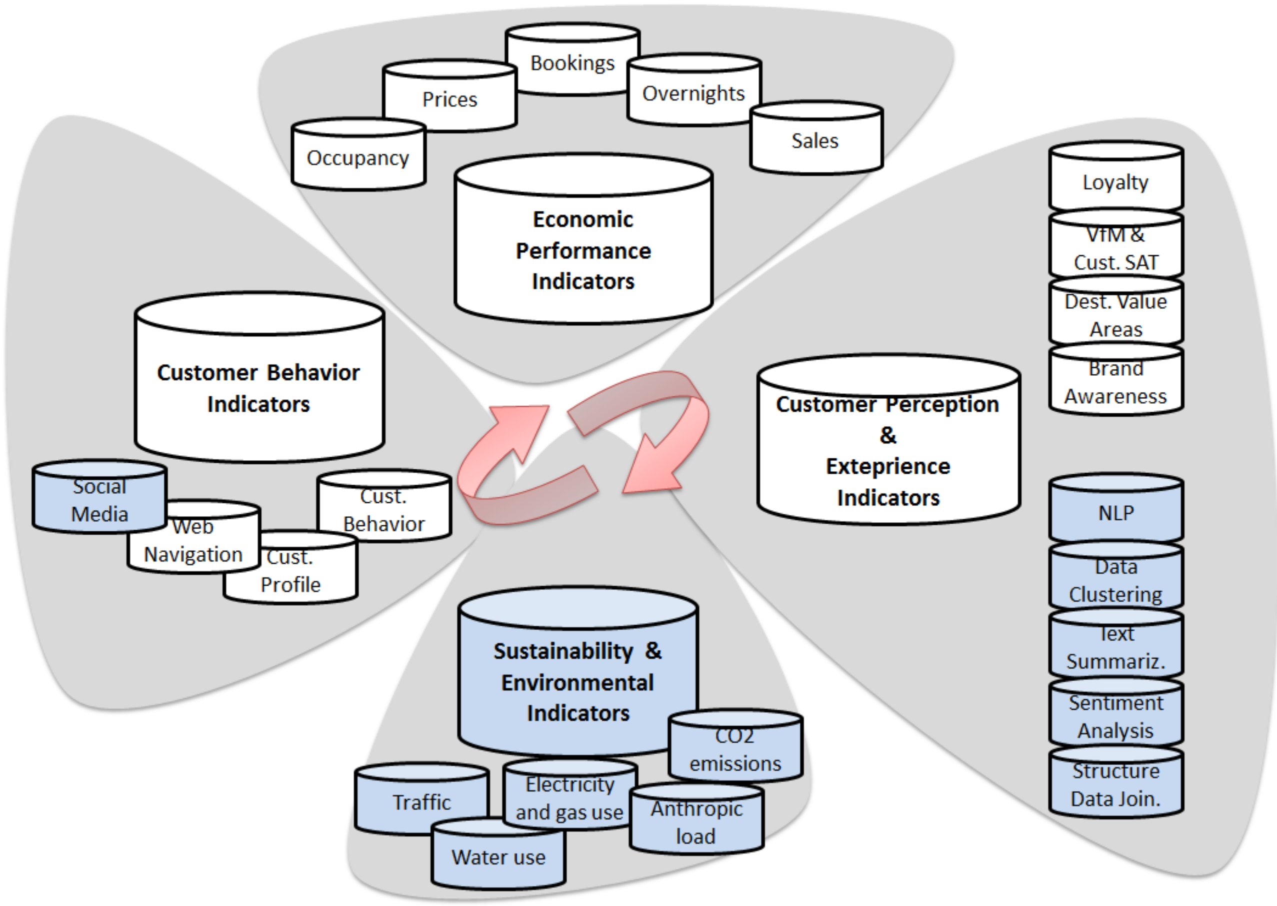

Fig. 2 Prototype of a DBIU

Moreover, our DBIU improves the “Functional, emotional value and satisfaction data” helping to enhance customer perception and experience indicators. The right bottom part of the figure shows (in blue) our improvements. The primary goal is to analyze 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 behavior 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, analyzer, and visualizer modules. For a detailed description see Mariani, Di Felice & Mura, 2016; 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 behavior and customer perceptions in real time with local SMES operating in the accommodation, and transportation industry. 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 continue 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, 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 up new opportunities in terms of modelling organisational operations and managing the firm in real time and has attracted interest on 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 and practically (Nudurupati et al., 2011; Nudurupati et al., 2016). Indeed, several questions arise. A key issue concerns the analysis of data availability and sources (Zhang, Yang and Appelbaum, 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 relations between BPA and ICT and its effect on their abilities to collect, select, manage and interpret data. Specifically, it highlights the key issues which 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, 2007, p. 7) and supporting effective PMSs design and adoption (Silvi et al., 2012). Examples are: decision support systems, expert 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 v*)* the visualization of business performance and dynamics and their management.

Figure 3 shows an example of 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 drilled 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.

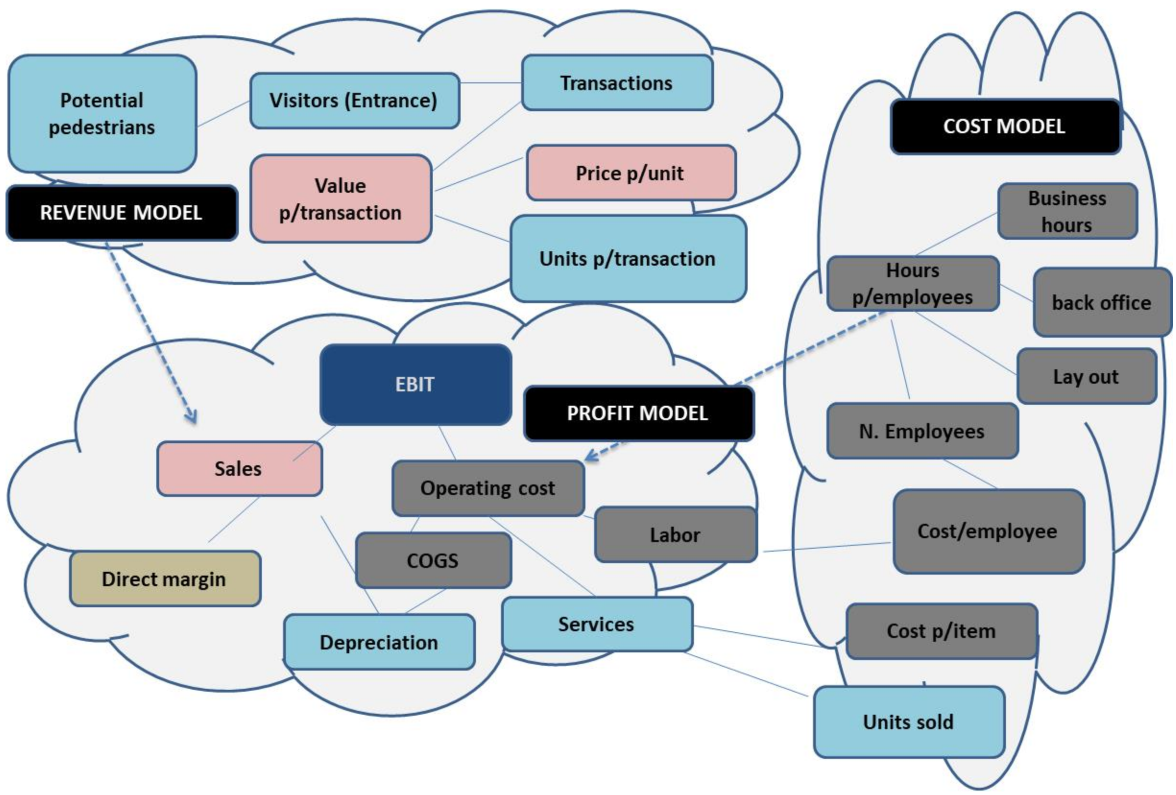

Fig. 3: Bookstore performance map

On the other hand, this performance and measurement system requires data availability, data analysis, data visualization technologies, analytical methods, routines and performance management skills, 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, these are: i) data collection, ii) data analysis, and iii) data visualization.

Data collection. Data is originated from different internal and external sources, stored in several systems, with different languages and forms (conversational, video, text, etc.), timing, size, accuracy, 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, Yang and Appelbaum (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 (Kimball and Ross, 2011), (Han and Kamber, 2011). Automatic reasoning and decision-making on 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 the 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. Since digital skills and talents are scarce, they represent an opportunity for research and education and value for community wellbeing.

Conclusions and implications for entrepreneurship research

Scholars of entrepreneurship cover a rather wide range of problems, contexts and processes, usually combining different social science perspectives. 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 have been inadequate on many grounds (too general, incomplete, self-reported etc. etc.). Data gathering, storage and manipulation have therefore become a key element in any research program, with often 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 on many different types of research and approaches as the various cases sketched in this chapter have tried to illustrate. We are at the break of a new dawn for reconsidering field data in a completely new light. 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 behavior and their implications for entrepreneurship. Case 1 offered us a specific example associated with the analysis of interpersonal networks. Second, the possibility to standardize 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 progresses in these areas showing clear opportunities, as Case 4 exemplified in the field of tourism. 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 the users with both the historical accounts and the most recent evidence. Case 2 as well as Case 3 discussed different examples related to datasets of different size, composition and span, ranging from research driven, to institutionally driven and to commercially driven ones. Fourth, and probably more evident in its short-term impact, decision making processes, tools and roles is 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.

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 ones can be added. We are looking forward to reading from other scholars theirs and we hope we have offered some inspirations to begin an exciting and unpredictable new journey.

References

Amaro, S., Duarte, P. and Henriques, C., 2016. Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, *59*, pp.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., Bastardoz, N., Liu, Y., Schriesheim, C. A. (2014). What makes articles highly cited? The Leadership Quarterly, 25 (1): 152-179. DOI: https://doi.org/10.1016/j.leaqua.2013.10.014

Arrow, T., and Kasberger, S. 2017. Introduction to Content Mine: Tools for Mining Scholarly and Research Literature. Virginia Tech. University Libraries. Handle: http://hdl.handle.net/10919/77525

Beaubien, L. 2012. Technology, change, and management control: a temporal perspective. *Accounting, Auditing & Accountability Journal*, 26(1), pp. 48-74.
Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The semantic web. *Scientific American*, *284*(5), pp. 28-37.
Bhimani, A. and Willcocks, L. 2014. Digitisation, "Big Data" and the transformation of accounting information. *Accounting and Business Research*, 44(4), pp. 469-490.
Cambria, E. and White, B. 2014. Jumping NLP curves: a review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, 9(2), pp. 48-57.
Cartwright, N. (2014). Causal Inference. In N. Cartwright & E. Montuschi (Eds.), Philosophy of Social Science: A New Introduction. OUP Oxford.
Cartwright, N., & Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), Philosophy of Social Science: A New Introduction. OUP Oxford
Chaabani, Y., Toujani, R. and Akaichi, J., 2017, June. Sentiment Analysis Method for Tracking Touristics Reviews in Social Media Network. In *International Conference on Intelligent Interactive Multimedia Systems and Services* (pp. 299-310). Springer, Cham.
CIMA, 2014. *Big Data. Readying business for the big data revolution.* [online] Available at: http://www.cgma.org/Resources/Reports/DownloadableDocuments/CGMA-briefing-big-data.pdf [Accessed 15 July 2015].
Cyganiak, R., Wood, D., and Lanthaler, M. 2014. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014. http://www.w3.org/TR/rdf11-concepts/ (last visited July 20 2017)
Davenport, T. H. and Harris, J. G. 2007. *Competing on Analytics*. Boston: The New Science of Winning. Harvard Business Press.
Davenport, T.H., Harris, J.G. and Morison, R., 2010. Analytics at Work: Smarter Decisions. Better Results, Boston: Harvard Business Press.
Dechow, N., Granlund, M. and Mouritsen, J. 2007. Interactions between information technology and management control. In: D. Northcott et al., eds., *Issues in Management Accounting*. 3rd ed. London: Pearson, Chapter 3.
Di Iorio, A., Nuzzolese, A. G., Peroni, S. (2013). Characterising Citations in Scholarly Documents: The CiTalO Framework. ESWC (Satellite Events) 2013: 66-77. DOI: 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. DOI: https://doi.org/10.1002/asi.22806
Droll, A., Shahzad, K., Ehsanullah, E., and 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), pp. 25-37.
Etzkowitz, H. and 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), pp.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. DOI: https://doi.org/10.1371/journal.pone.0049476
Ferrucci, D., Lally, A., Verspoor, K., and Nyberg, E. 2009. Unstructured Information Management Architecture (UIMA) Version 1.0. OASIS Standard. http://docs.oasis-open.org/uima/v1.0/uima-v1.0.html (last visited 29 July 2017)

Fini, R., Fu, K., Mathisen, M.T., Rasmussen, E., and Wright, M. 2017. Institutional determinants of university spin-off quantity and quality: a longitudinal, multilevel, cross-country study. *Small Business Economics*, 48(2), pp. 361-391.

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), pp. 637-659.

Fink, J. L., Fernicola, P., Chandran, R., Parastatidis, S., Wade, A., Naim, O., Quinn, G. B., and Bourne, P. E. (2010). Word add-in for ontology recognition: semantic enrichment of scientific literature. BMC Bioinformatics, 2010 (11): 103 DOI: https://doi.org/10.1186/1471-2105-11-103

Fuchs, M., Höpken, W., and Lexhagen, M., 2014. Big data analytics for knowledge generation in tourism destinations – A case from Sweden, *Journal of Destination Marketing & Management*, 3(4), pp. 198-209.

George, G., Osinga, E.C., Lavie, D., and Scott, B.A. 2016. Big data and data science methods for management research. *Academy of Management Journal*, 59(5), pp. 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. and Kamber, M. 2011. *Data mining: concepts and techniques*. 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., and Seaborne, A. 2013. SPARQL 1.1 Query Language. W3C Recommendation 21 March 2013. http://www.w3.org/TR/sparql11-query/ (last visited 28 July 2017)

IFAC 2011. *Predictive Business Analytics: Improving Business Performance with Forward-looking Measures*. [online] Available at: https://www.ifac.org/publications-resources/predictive-business-analytics-improving-business-performance-forward-looking- [Accessed 26 July 2017].

Kimball, R. and Ross, M. 2011. *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.

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., McKibbon, 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. DOI: https://doi.org/10.1136/bmj.39482.526713.BE

March, J.G. and Simon, H.A., 1958. Organizations.

Mariani, M. M., Di Felice, M., and Mura, M., 2017. The determinants of Facebook social engagement for National Tourism Organisations' Facebook pages: A quantitative approach. *Journal of Destination Marketing & Management*, forthcoming.

Mariani, M.M, and 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), pp. 89-109.

Mariani, M.M., Baggio, R., Fuchs, M. and Höpken, W., 2017. Business Intelligence and Big Data in Hospitality and Tourism: A Systematic Literature Review. *International Journal of Contemporary Hospitality Management*, forthcoming.

Mariani, M.M., Di Felice, M. and Mura, M., 2016. Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, *54*, pp.321-343.
McAfee, A., and Brynjolfsson, E. 2012. Big data: The management revolution. *Harvard Business Review*, 90, pp. 61–67.
McPhee, C. 2016. Editorial: Managing Innovation. *Technology Innovation Management Review*, 6(4), pp. 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 (Second Edition). W3C Recommendation 11 December 2012. http://www.w3.org/TR/owl2-syntax/ (last visited 28 July 2017)
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. and Hardman, J. 2016. Contemporary performance measurement and management (PMM) in digital economies. *Production Planning & Control*, 27(3), pp. 226-235.
Nudurupaty, S.S., Bititci, U.S., Kumar, V. and Chan, F.T.S. 2011. State of the art literature review on performance measurement. *Computers and Industrial Engineering*, 60, pp. 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. DOI: https://doi.org/10.1002/asi.23209
Opthof, 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. DOI: https://doi.org/10.1016/S0008-6363(02)00712-5
Orwell, G. (1949). *1984*. Houghton Mifflin Harcourt.
Parker, G.G., Van Alstyne, M.W. and Choudary, S.P., 2016. *Platform revolution: How networked markets are transforming the economy--and how to make them work for you*. WW Norton & Company.
Perkmann, M., Neely, A. & Walsh, K. 2011b. How should firms evaluate success in university-industry alliances? A performance measurement system. *R&D Management*. 41 (2), 202-216.
Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A. and Sobrero, M. 2013. Academic engagement and commercialisation: A review of the literature on university-industry relations*. Research Policy*. 42 (2), 423-442.
Peroni. S., Dutton, A., Gray, T., and Shotton D. 2015. Setting our bibliographic references free: towards open citation data. *Journal of Documentation*, 71(2), pp. 253-277.
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. DOI: https://doi.org/10.1108/JD-12-2013-0166
Polanyi, M. 2013. *The tacit dimension.* Garden City, N.Y.: 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.

Ransbotham, S., Kiron, D. and 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] Available at: http://sloanreview.mit.edu/projects/the-hard-work-behind-data-analytics-strategy/ [Accessed 14 July 2016.

Ritter, T., & Gemunden, G. 2003. Interorganizational relationships and networks: An overview. *Journal of Business Research*, 56 (9), 691-697

Sateli, B., and Witte, R. (2015). Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud. PeerJ Computer Science: e37. DOI: https://doi.org/10.7717/peerj-cs.37

Shotton, D. 2009. Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, 22(2), pp. 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. and Visani, F. 2012. Business Performance Analytics: Level of Adoption and Support Provided to Performance Measurement Systems. *Management Control*, 3 (Special Issue), pp. 117-142.

Simon, H. A. 1972. Theories of bounded rationality. *Decision and organization* 1, no. 1: 161-176.

Sundararajan, A., 2016. *The sharing economy: The end of employment and the rise of crowd-based capitalism*. 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. DOI: 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), pp. 169-183.

van Knippenberg, D., Dahlander, L., Haas, M., and George, G. 2015. Information, attention, and decision-making. *Academy of Management Journal*, 58(3), pp. 649-657.

Vrandečić, D., and Krötzsch M. 2014. Wikidata: a free collaborative knowledgebase. *Communication of the ACM*, 57(10), pp. 78-85.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive Measures*. SAGE Publications

West, J., Vanhaverbeke, W. & Chesbough, H. 2006. Open innovation: a research agenda. In: Chesbrough, H., Vanhaverbeke, W. & West, J. ed. *Open innovation: Researching a new paradigm*. Oxford: Oxford University Press, pp. 285-307.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 3*. DOI: https://doi.org/10.1038/sdata.2016.18

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), pp. 1188-1198.

Williamson, O.E., 1979. Transaction-cost economics: the governance of contractual relations. *The journal of Law and Economics*, *22*(2), pp.233-261.

Zhang, C. 2015. DeepDive: A Data Management System for Automatic Knowledge Base Construction. Ph.D. Dissertation, University of Wisconsin-Madison. http://cs.stanford.edu/people/czhang/zhang.thesis.pdf (last visited 29 July 2017)

Zhang, J., Yang, X. and Appelbaum, D. 2015. Toward Effective Big Data Analysis in Continuous Auditing. *Accounting Horizons*, 29(2), pp. 469-476.

Zikopoulos, P., and Eaton, C. 2011. *Understanding big data: Analytics for enterprise class Hadoop and streaming data*. New York, NY: McGraw-Hill