Big and open linked data analytics ecosystem: Theoretical background and essential elements

Martin Lněnička and Jitka Komárková

Abstract: *Big and open linked data are often mentioned together because storing, processing, and publishing large amounts of these data play an increasingly important role in today's society. However, although this topic is described from the political, economic, and social points of view, a technical dimension, which is represented by big data analytics, is insufficient. The aim of this review article was to provide a theoretical background of big and open linked data analytics ecosystem and its essential elements. First, the key terms were introduced including related dimensions. Then, the key lifecycle phases were defined and involved stakeholders were identified. Finally, a conceptual framework was proposed. In contrast to previous research, the new ecosystem is formed by interactions of stakeholders in the following dimensions and their sub-dimensions: transparency, engagement, legal, technical, social, and economic. These relationships are characterized by the most important requisites and public policy choices affecting the data analytics ecosystem together with the key phases and activities of the data analytics lifecycle. The findings should contribute to relevant initiatives, strategies, and policies and their effective implementation.*

Key words: *big and open linked data; ecosystem approach; dimensions; data analytics lifecycle; stakeholders; conceptual framework*

1 Introduction

The efforts of governments to be more transparent and responsive to citizens' demands have been increasing in the last few years (Attard et al., 2015; Ruijer et al., 2017). In response to this emerging pressure, the term citizen engagement has undergone a major transformation from a passive consumer to an active prosumer (Millette and Hosein, 2016; Susha et al., 2015). This term underpins participation, collaboration, and cooperation in policy-making and governance through Information and Communication Technologies (ICT) and various communication and delivery channels for open and linked data (Lněnička et al., 2016). At the same time, due to the rapid pace of technological change, the amount of digital data available is growing fast and these big data are inevitably evolved into a Big and Open Linked Data (BOLD) concept (Dwivedi et al., 2017; Janssen and Kuk, 2016; Matheus and Janssen, 2015; Saxena, 2017).

Taking a closer view on each of these terms, big data are characterised by their volume (large amounts of data available), variety (different types of data occurring in datasets), and velocity (speed at which data are generated), together with the availability of new platforms, tools and techniques to process data (de Mauro et al., 2015; Lake and Drake, 2014; Tsai et al., 2015). Open data are represented by their free availability and accessibility with no legal restrictions imposed on their reuse (Janssen et al., 2012). Linked data are driven by the effort to interlink all the available datasets into a single space of global data-centric economy in which involved stakeholders can reuse datasets and create value from them. Data analytics is then a means to strengthen the data-centric economy. However, research integrating these concepts together has only recently begun to emerge and need to be explored (Bertot et al., 2014; Janssen and Kuk, 2016; Matheus and Janssen, 2015; Saxena, 2017; Weerakkody et al., 2017).

In a well-functioning democratic society, citizens need to be informed and have access to information on government policies and progress (World Wide Web Foundation, 2015). According to Weerakkody et al. (2017), public agencies and institutions are becoming increasingly transparent with their data to establish the whole new paradigm of big open data. Hardy and Maurushat (2017) argued that a major purpose of releasing these data is to drive innovation through big data analytics. Therefore, in light of recent events in open government movement that are intended to increase transparency through big data analytics, it is becoming difficult to ignore the existence of a variety of emerging analytics practices and various forms of analytical platforms, tools, and services. Those may offer a wide range of benefits to cover the whole lifecycle of public data and help the related efforts of governments to success (Janssen and Kuk, 2016; Khan et al., 2017; Lněnička and Komárková, 2015; Peng et al., 2016; Vossen, 2014). Along similar lines, Millette and Hosein (2016) state that: "*In the age of big data processing, the power of the open data resource has yet to be fully harnessed to the betterment of human-kind.*"

The basic idea behind BOLD analytics is that open datasets can be of any size. Many of them consist of big datasets that are extremely large and/or complex, offering the possibilities of identifying previously impossible levels of insights, granularity of analysis, and relationships between elements in the dataset as linked data (Attard et al., 2016; Bertot et al., 2014; Janssen and Kuk, 2016; Máchová and Lněnička, 2017). Dwivedi et al. (2017) stated that the BOLD concept is a fledgling and rapidly evolving field that creates new opportunities for innovation. It is defined by Janssen and Kuk (2016) as "*the integration of diverse data, without predefined restrictions or conditions of use, to create new insights.*" As noted by Bertot et al. (2014) and Janssen and van den Hoven (2015), BOLD results in new opportunities and have the potential to transform e-government services and the interaction between governments, citizens, and the business sector. The open data movement is foundational for BOLD to succeed, as it ensures publicly accessible datasets through analytics processes (Janssen and Kuk, 2016). Kalampokis et al. (2013) claimed that the real value of BOLD will unveil from performing data analytics on top of combined datasets that were previously closed in disparate sources and can be linked to provide unexplored insights. The value arises from the ability to work with these data to develop actionable information (Kaisler et al., 2013; Zuiderwijk et al., 2014, Khan et al., 2017).

On the other hand, Bertot and Choi (2013) reported that there is a need for a BOLD governance model to more fully address the policies and practices surrounding these data. According to Kaisler et al. (2013), it is important to design appropriate systems to handle these data effectively and analyse them to extract relevant meaning for decision-making. In addition, the demand for quicker and high-quality decisions based on these data requires: (1) new forms of participation to obtain data from various sources, (2) new ways to process data, including algorithms that are capable of dealing with the vast amount of data, and (3) redesigned processes to include more people in the interpretation of results (Höchtl et al., 2016). As a result, BOLD analytics as a concept is wider due to the emphasis on related ICT, knowledge, usage skills, and technical requirements (Sayogo et al., 2014; Lněnička and Komárková, 2015). In this regard, Saxena (2017) presented a trajectory for BOLD wherein progressive stages for BOLD and their elements have been identified.

However, most authors make no attempt to differentiate between different phases and activities of the BOLD analytics lifecycle, requirements, and responsibilities of involved stakeholders, or dimensions that are affected by these data. There is also lacking any connection between these elements within an ecosystem and its perspective in the public sector context.

Regarding these shortcomings, an ecosystem approach was applied to identify these elements. The use of this approach should ensure that the public agencies and institutions will fulfil their tasks with quality, efficiency, and stakeholders' satisfaction, as required by initiatives of open government movement (Bogdanović-Dinić et al., 2014; Dawes et al., 2016; Heimstädt et al., 2014; Martin et al., 2017; Reggi and Dawes, 2016; Zuiderwijk et al., 2014). More precisely, Dawes et al. (2016) emphasized this approach due to its "*emphasis on an evolving, self-organizing system of feedback and adjustment among actors and processes*". Therefore, in contrast to previous research papers and review articles, the concept of the BOLD analytics ecosystem has never been formally proposed or clearly delineated in such a way as to take into account all the essential elements. The ecosystem perspective also provided a framework within which most other aspects represented by big/open/linked data can be incorporated.

This article is structured as follows. This introduction is followed by Section 2 explaining the research methodology. Subsequently, background and the key concepts of this article are defined in Section 3. The following Section 4 provides essential elements of the BOLD analytics ecosystem, including dimensions, lifecycle, stakeholders, and analytics framework. Discussion and limitations are presented afterwards in Section 5. Finally, conclusions are made in Section 6 by addressing the contributions of the article and issues for further research are outlined.

2 Research Methodology

A research methodology in this article has been developed in order to provide a theoretical background and essential elements of BOLD analytics. In this regard, it aims to answer the following research questions: (1) Which elements of BOLD analytics are found in the literature? (2) How these elements shape the ecosystem of existing concepts and their relationships? (3) How are these elements and their relationships represented in a conceptual framework and to what extent is it reinforcing of existing frameworks?

For this purpose, three concepts of current data analytics perspectives were used as a lens, namely big, open, and linked data. Since it is now becoming more than obvious that data openness is complementary to the shift to big data solutions and the need to link the data from different sources, we are convinced that these perspectives should be merged together to reflect new requirements on the data analytics ecosystem. By reshaping and forming the BOLD analytics ecosystem, we are able to provide a new point of view on data processing and management regarding the use of BOLD in the public sector context.

In order to gather background for reshaping and validating the BOLD analytics ecosystem, a comprehensive literature review was conducted to identify research papers and review articles relevant to this topic. It looked for elements and factors affecting each of data analytics perspective in relevant papers published in related journals and conferences during the last decade. From the comprehensive literature review, we are able to identify dimensions of the BOLD analytics ecosystem, BOLD analytics lifecycle phases, stakeholders in the BOLD analytics ecosystem, and define BOLD analytics framework. Our aim was to look for elements that are characteristic for BOLD and at the same time are missing in the existing frameworks.

The comprehensive literature review was done by following the approach proposed by Levy and Ellis (2006). The following steps were followed: a) a systematic collection of input papers for the review; b) a systematic processing of relevant papers; c) a systematic synthesis of outputs. The following electronic databases were searched during the collection of input papers: Web of Science and Scopus. Although we considered other electronic databases to be used for this systematic review, we decided against including it, since their content is available through the listed electronic libraries, thus making the use of these databases redundant. Input papers were searched within the time period 2003 – 2018. Both reviewed journals (including articles in press) and conference proceedings were included. Combinations of the following key words were used for search: big data analytics, open linked data analytics, open data analytics, elements, dimensions, lifecycle, stakeholders, ecosystem, and framework.

The precise list of combinations is presented in Table 1. It resulted in 2 167 candidate papers and articles. There were many overlapping papers and articles included as a result of searches by different key words and by different search engines. Table 1 shows the total numbers including duplicate contributions. Mendeley was used to manage the literature review process. It allowed authors to exclude duplicate contributions. Next, the candidate contributions were reviewed by the authors according to the title and key words to exclude not relevant contributions. The term "framework" resulted in the highest number of not relevant papers. It should be also noted that although big data analytics is the most dominant concept in this field, more than one-third of the relevant papers contain both or all three data analytics perspectives. Most of them were published after 2015. Then, the exclusion and inclusion criteria were applied in order to reduce the bias in the results. The following criteria were used:

1. We considered the papers with title, abstract, and key words in English.
2. Papers that were not accessible online were excluded.
3. Only papers of at least three pages were taken into account.
4. We included papers that regarded elements of one or more data analytics perspectives as their main focus.
5. The papers that described only case studies with no implications for the public sector were excluded.

These papers were analysed according to the title, abstract and key words. From the remaining 173 papers, those relating to our research questions were selected. Then, global reports published by international organizations such as United Nations were added to the list of papers to understand global trends. Finally, we arrived at a selection of 84 papers. These papers were decomposed and the main elements were identified through the systematic synthesis. All elements were grouped and classified regarding their most common characteristics to understand better the overall picture of the BOLD analytics ecosystem. In the final stage, the conceptual framework was proposed including relationships between elements.

Table 1: Results of the systematic search

Key words combinations	**Web of Science**		**Scopus**		
	proceed-ings	journals	proceed-ings	journals	book chapters
"big data analytics" AND elements	26	32	47	33	4
"big data analytics" AND dimensions	47	43	65	48	6
"big data analytics" AND lifecycle	10	14	16	20	2
"big data analytics" AND stakeholders	20	18	26	29	2
"big data analytics" AND ecosystem	44	24	66	33	8
"big data analytics" AND framework	370	221	512	296	47
open linked "data analytics" AND elements	2	1	0	0	0
open linked "data analytics" AND dimensions	1	0	2	0	0
open linked "data analytics" AND lifecycle	2	0	4	1	0
open linked "data analytics" AND stakeholders	0	1	1	1	0
open linked "data analytics" AND ecosystem	1	0	0	1	0
open linked "data analytics" AND framework	8	1	8	1	0
"open data analytics" AND elements	0	0	0	0	0
"open data analytics" AND dimensions	0	0	0	0	0
"open data analytics" AND lifecycle	0	0	0	0	0
"open data analytics" AND stakeholders	0	0	0	0	0
"open data analytics" AND ecosystem	0	0	0	0	0

“open data analytics” AND framework	1	0	0	1	0
Sum	**532**	**355**	**747**	**464**	**69**

Table 2 shows cited top 10 journals from Web of Science (all Q1 or Q2), top 3 journals from Scopus, and top conferences, including a number of papers.

Table 2: Top cited journals and conferences

Journals	**Impact Factor**	**Number**
Government Information Quarterly	4.009	8
Information Systems Frontiers	3.232	2
Journal of Strategic Information Systems	4.313	1
Information Processing & Management	3.444	1
Computers & Industrial Engineering	3.195	1
Journal of Biomedical Informatics	2.882	1
Safety Science	2.835	1
VLDB Journal	2.689	1
IEEE Intelligent Systems	2.596	1
Mobile Networks and Applications	2.497	1
Information Polity	Scopus: 2.17	3
Journal of Big Data	Scopus: 8.51	2
Transforming Government: People, Process and Policy	Scopus: 2.39	1
Conferences	**Indexed in**	**Number**
Hawaii International Conference on System Sciences	Web of Science	4
IFIP WG 8.5 International Conference: EGOV	Web of Science	3
International Conference on Digital Government Research	Scopus	1
International Conference on Electronic Government and the Information Systems Perspective, EGOVIS	Scopus	1

Table 3 shows the structure of papers from these journals according to particular categories of journals. Papers are not summarized because several journals cover more categories.

Table 3: The structure of the categories of the top cited journals

Category	**Web of Science**	**Scopus**
Computer Science: Artificial Intelligence	1	
Computer Science: Computer Science Applications		1
Computer Science: Hardware & Architecture	2	
Computer Science: Information Systems	6	5
Computer Science: Interdisciplinary Applications	3	
Computer Science: Theory & Methods	2	
Decision Sciences: Information Systems and Management		3
Engineering: Electrical & Electronic	1	
Engineering: Industrial	3	
Information Science & Library Science	10	
Management	1	
Operations Research & Management Science	1	
Social Sciences: Communication		3
Social Sciences: Public Administration		4
Social Sciences: Sociology and Political Science		3
Telecommunications	1	

3 Background and Concepts

In addressing the literature review and background, three concepts of current data analytics perspectives that are associated with this article were examined. First, each perspective was defined in its own section. Then, a data analytics concept was introduced as an umbrella term for them.

3.1 Big Data

At present, although the importance of big data has been generally recognized, people still have different opinions on the definition of this term (Chen et al., 2014; Gandomi and Haider, 2015; Tsai et al., 2015). As reported by Gandomi and Haider (2015) and Klievink et al. (2017), it is mostly due to fast advances in technology, exactly what can be considered big data is always changing, making it hard to express in specific and measurable terms. Thus, definitions of big data also depend upon the industry. In general, big data represent the datasets that could not be perceived, acquired, managed, and processed by traditional ICT and software/hardware tools within a tolerable time (Chen et al., 2014). The rise of big data is enabled by recent increases storage and processing capacity as well as increases in the number of devices collecting and sharing data. Moreover, big data span a range of data types such as text, numeric, image, video, or combinations thereof, and they can cross multiple data platforms (Bertot and Choi, 2013; Gandomi and Haider, 2015). Finally, although the technology perspective, of big data is the most important, other perspectives such as business, social, and environmental have to be taken into account while discussing big data (Lake and Drake, 2014).

The result of an extensive literature review on big data definitions by de Mauro et al. (2015) concluded that a consensual definition of big data would be: "*Big data represent the information assets characterized by such a high volume, velocity, and variety to require specific technology and analytical methods for their transformation into value.*" In addition to these 3V's (volume, velocity, variety), big data are often defined through their variability, veracity, and value (Lake and Drake, 2014). Also important is the fact that these dimensions are not independent of each other. As stated by Gandomi and Haider (2015), if one dimension changes, the likelihood increases that another dimension will also change as a result. As summarized by Klievink et al. (2017), big data are characterized by: use and combining of multiple, large datasets, from various sources, external and internal to the organization; use of incoming data streams in real-time or near real-time; development and application of advanced analytics and algorithms, distributed computing and/or advanced technology to handle very large and complex computing tasks; innovative use of existing datasets and/or data sources for new and different platforms, tools, and services. Therefore, big data analytics requires unique platforms, tools, and services that reduce time and can offer distributed and scalable solutions, such as those included in the Apache Hadoop ecosystem (Lněnička and Komárková, 2015; Vossen, 2014).

3.2 Open Data

Open data are based on the idea that certain kinds of data should exist beyond the limits of copyright, patents, or censorship (Janssen et al., 2012). Countries committed to open government and open data believe that openness will, among other things, promote transparency, fight corruption, energize civic engagement, and facilitate the creation of new services that deliver social and commercial value (Attard et al., 2015; Ruijer et al., 2017; Sayogo et al., 2014). Data are considered open when they are freely available and shareable online to reuse, republish, and transform them into new structures, without charge (World Wide Web Foundation, 2015). It is important to note the distinction between public data and open data. While public data are made freely available to the general public, they are not necessarily open. Open data have a particular license of use and distribution (Attard et al., 2015). However, before transparency or any of the other effects can happen, public data have to be disclosed in the first place (Barry and Bannister, 2014).

Although there are many different sources of open data, government data are particularly important because of their scale, breadth, and status as the main source of public sector information on a wide range of subjects (Kučera et al, 2013; Van der Waal et al., 2014). These are a subset of open data, and are simply government-related data that are made open to the public (Attard et al., 2015). Since governments collect big data from multiple sources (United Nations, 2016), the more government data that is available as open data, the greater the opportunities are for the stakeholders to reuse them (Hardy and Maurushat, 2017). On the other hand, not all government data can be published as open data. For example access to some datasets is restricted for national security or privacy reasons and thus it cannot be made publicly available for reuse (Dwivedi et al., 2017; Kučera et al., 2013).

A more specific definition was given by Geiger and von Lucke (2012) asserting that open government data are "*all stored data of the public sector which could be made accessible by government in a public interest without any restrictions for usage and distribution*." Janssen et al. (2012) then added the aspect of funding sources and define them as "*non-privacy-restricted and non-confidential data which are produced with public money and are made available without any restrictions on its usage or distribution.*" Sayogo et al. (2014) conceptualized the definition of open government data loosely to indicate open publication of data collected and stored by government agencies, which is consented by laws to be made accessible to the general public through a single data portal. Open data help to reduce the time and money that stakeholders need to invest to understand what government is doing and to hold it to account. Moreover, comparing and combining data from different sources become faster and easier, even across national boundaries. This enhances the ability of stakeholders to find solutions to complex development problems and create new products and services from these data (Janssen et al., 2012; Jetzek et al., 2014; Kucera and Chlapek, 2014; World Wide Web Foundation, 2015). A summarization of the basic principles that permeate the open government idea can be found, e.g., in Veljković et al. (2014). These are important since they distinguish between transparency and engagement dimensions.

3.3 Linked Data

Linked data and semantic web technologies have the potential for solving many challenges in open government data, as well as possibly lowering the costs and complexity of developing government data-based applications (Attard et al., 2016). Linked data are mostly taken together with open data as Linked Open Data (LOD) because when open data are interlinked to provide more context, greater opportunities for stakeholders to exploit the data for innovative purposes are provided (Dwivedi et al., 2017; Geiger and von Lucke, 2012). Geiger and von Lucke (2012) also defined LOD as: "*all stored data connected via the World Wide Web which could be made accessible in a public interest without any restrictions for usage and distribution*." The concept of linked data enables for huge amounts of data to be processed and analysed efficiently.

Linked data refer to a set of best practices for publishing and connecting unstructured data to become structured. It allows data providers to link their data with other different sources over the Web using a standard mechanism for specifying the existence and meaning of connections between items described in these data (Heath and Bizer, 2011). The increased rate of adoption of linked data best practices has led the Web to evolve into a global data space containing billions of assertions, i.e. the Web of Data, where both documents and data are linked (Heath and Bizer, 2011; Van der Waal et al., 2014). The evolution of the Web of Data enabled to combine distributed datasets, explore relationships between them, and support development of new applications and services (Attard et al., 2015; Janssen and Kuk, 2016). The concept of linked data has provided access to more data and has enabled to process them automatically (Heath and Bizer, 2011; Janssen and Kuk, 2016).

3.4 Data Analytics

The importance of data analytics lies in efficient processing of these data to convert into highly valued insights needed to drive decision making (Gandomi and Haider, 2015). According to Cárdenas et al. (2013), big data analytics is "*the large-scale analysis and processing of information*". It has become an active part of data analyses in various fields because large volumes of data are recorded or submitted every minute.

Analyses of security data can be given as an example. Various systems, e.g. operating systems, web servers, and intrusion detection systems log activities of users, system events and network flows. Their short-term analysis has been a typical task for many years. Big data analytics introduced possibility of long-term and large-scale analyses, including unstructured data sources, which can e.g. lead to frauds detection or definition of safety rules and establishing of more defence mechanisms (Ouyang et al., 2018; Lněnička et al., 2017). Tiwari et al. (2018) investigated big data analytics in supply chain management between 2010 and 2016. They clearly showed increasing importance of big data analytics for industries and increasing need for suitable approaches. Günther et al. (2017) conducted another literature review to point out importance of big data analytics to get a real value from contemporary data. Analytics techniques for text, audio, video, and social media data, as well as predictive analytics were reviewed by Gandomi and Haider (2015).

Data analytics is becoming reality in governments due the quantity of data available and the evolution of technologies and techniques used to analyse data (Chatfield and Reddick, 2018; Khan et al., 2017; Matheus and Janssen, 2015; Tsai et al., 2015). The idea is that by releasing large government datasets, stakeholders will be able to draw new insights from these data and contribute innovative solutions to complex policy problems (Hardy and Maurushat, 2017). Chen et al. (2014) presented three main fields of big data analytics in the public sector: scientific exploration, regulatory enforcement, and data as the basis of public information services. The applications of big data analytics in the policy-making process was discussed by Höchtl et al. (2016). They focused on its applicability for the large-scale interpretation of public opinion. Chatfield and Reddick (2018) focused on big data analytics-enabled customer agility and responsiveness in order identify key institutional mechanisms for linking them to public value creation. The concept of data analytics focusing on linked open government data was described by Kalampokis et al. (2013). They claimed that the technical infrastructure is essential for employing data analytics in a decentralized manner on the Web. Similarly, Debattista et al. (2015) explored the potential of data analytics in the Web of Data focusing on the parallels between methods used for big data and linked data.

However, performing analytics on BOLD is a challenging task considering the functional requirements resulting from the specific characteristics of these data. To deal with this, it is necessary to merge these requirements into a single framework and then build the right mix of platforms, tools, services, and analytics techniques upon the framework.

Since BOLD analytics requires the specific technologies, techniques, and also skills, the literature identifies many barriers that are hindering its adoption. Kaisler et al. (2013) as well as Khan et al. (2017) identified some of the major issues in data storage, management, and processing. Among risks, Janssen et al. (2012) addressed task complexity, use and participation, legislation, information quality, technical and institutional factors. In similar lines, Janssen and van den

Hoven (2015) claimed that without taking into account the underlying context, the act of generalization based on BOLD may lead to unreliable or wrong plans and conclusions. In addition, as reported by Hardy and Maurushat (2017), not all analysis of open data will involve big data analytics, but big data analytics rely heavily on open data. This supports the argument in favour of the existence of the BOLD analytics ecosystem.

Data quality is recognized as an important property of all types of data, either for large unstructured datasets or small structured datasets (Barry and Bannister, 2014; Chen et al., 2014; Günther et al., 2017; Máchová and Lněnička, 2017; Wang et al., 2015). On the basis of incomplete or incorrect information available in the form of BOLD, there are possibilities to make wrong decisions (Weerakkody et al., 2017). In the case of open government data available through open data portals, data provenance is mainly ensured by a history of a dataset's changes over time including associated metadata. On the other hand, big data pose challenges for provenance and quality (Chen et al., 2014; Wang et al., 2015). It is due to the existence of different components needed to perform analytics on these data, such as Hadoop, Spark, etc. (Bach et al., 2013; Lněnička and Komárková, 2015; Vossen, 2014). In this regard, Wang et al. (2015) introduced a reference architecture for big data provenance platform that enables to decide on components in each sub-system based on the targeted provenance usage scenario and capability.

A number of drivers that are helping overcome the barriers to BOLD analytics can be found in the literature. For example, Bertot and Choi (2013) recommended to: (1) review and recalibrate data policies, (2) build robust data platform and architecture, (3) provide significant computing power to process, analyse, manipulate, and represent these data through visualizations, (4) use data and metadata standards, (5) share data across sectors, and (6) foster research and data communities. In their follow-up paper, they extended this list by focusing on the data lifecycle phases (Bertot et al., 2014). Similarly, Dwivedi et al. (2017) found out that improvement in organisational factors and legal aspects lead to better access to data, superior awareness (including awareness of the platform where data are published), and data licensing. Finally, Weerakkody et al. (2017) used an extended technology acceptance model to empirically examine the factors affecting users' behavioural intentions towards public sector BOLD analytics. Their results suggested stakeholders were interested in using these data, if there is evidence of them being useful and more insightful in comparison to other data forms. Taking into account these findings, the BOLD analytics ecosystem is shaped around data lifecycle phases that will deliver data to stakeholders in an accessible, accurate, and actionable manner.

4 Essential Elements of the BOLD Analytics Ecosystem

Only little research has been performed on the essential dimensions of the big/open/linked data analytics ecosystems, while these ecosystems are important, as they may help in realizing the full value of these data (Dawes et al., 2016; Martin et al., 2017; Ubaldi, 2013; Zuiderwijk et al., 2014). As noted by Heimstädt et al. (2014), the existing literature contextualises digital ecosystems as cyclical, sustainable, demand-driven environments oriented around agents that are mutually interdependent in the delivery of value. However, with the increase of computational complexity powered by big data and requirements of open government movement, there is a need to define the BOLD analytics ecosystem, its dimensions, data lifecycle, and involved stakeholders.

The following dimensions were identified from the literature review. Each relevant paper was subjected to a more thorough evaluation in order to find the main field of focus and classify papers into appropriate dimensions. It should be noted that the number of dimensions and their elements resulted from this evaluation and were not previously defined. The supporting sources for each dimension are in Table 4. Each paper can be labelled with more than one dimension.

Table 4: Number of supporting papers for each dimension

Dimension	Number of papers
Transparency	21
Engagement	18
Technical	37
Social	16
Economic	29
Legal	15

4.1 Dimensions of the BOLD Analytics Ecosystem

The first step is to identify and interpret the key dimensions of BOLD analytics that influence its applicability. Their identification is crucial to specification what elements are relevant to the BOLD analytics ecosystem. In addition, it provides the linkage between theory and practice and clarifies the basic requirements for BOLD analytics in the public sector context. These dimensions and sub-dimensions are shown in Figure 1.

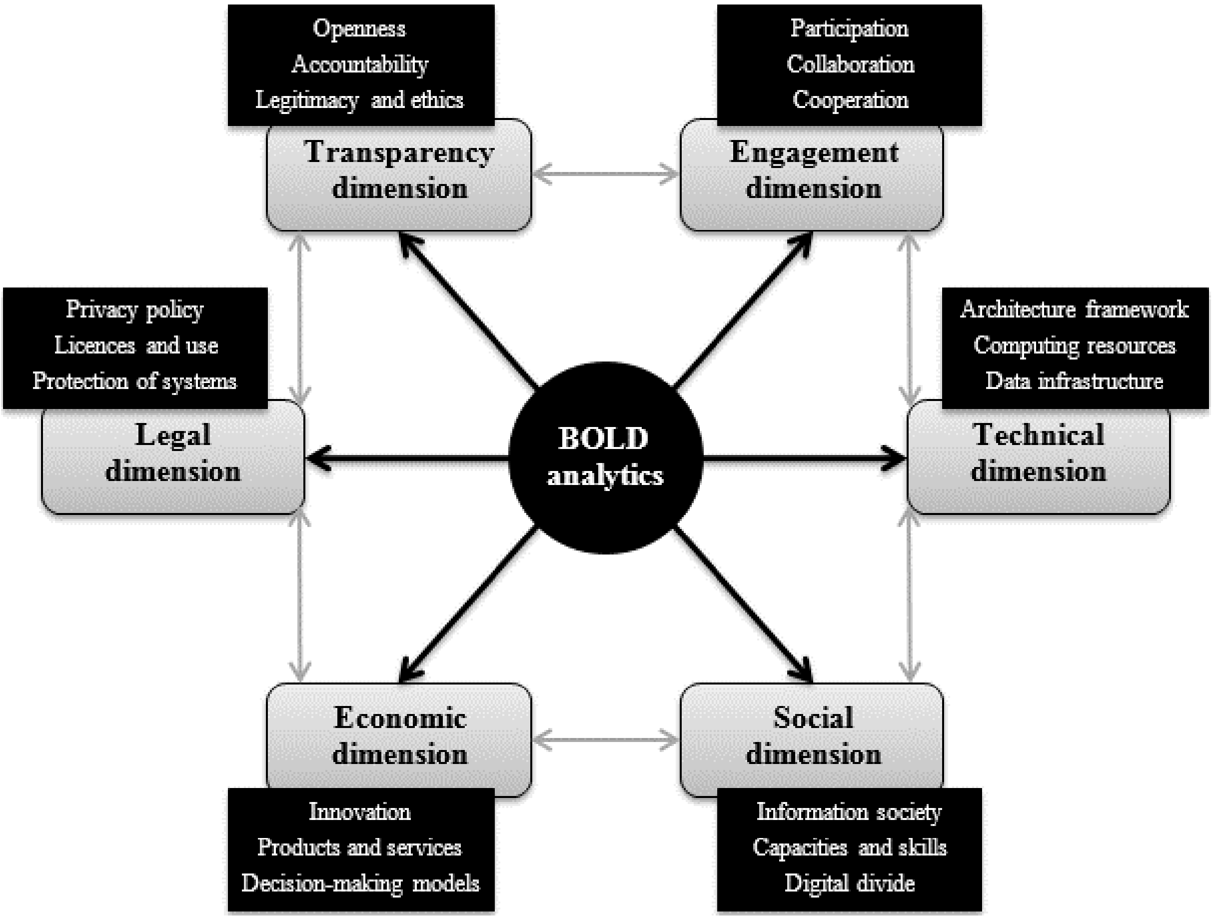

Figure 1: The main dimensions of the BOLD analytics ecosystem in the public sector context

Transparency Dimension

In the context of the public sector, the main goal of BOLD analytics is expressed by the transparency dimension. The willingness of public agencies and institutions to open up their data is generally managed by the effort of showing them in a better light. According to Matheus and Janssen (2015), transparency is based on the two major synonymous concepts used in literature, namely accountability and openness. In this article, this dimension is consisted of openness, accountability, and legitimacy and ethics aspects regarding the move towards a transparent governance process. A more detailed view on the concept of open data transparency is given by Reggi and Dawes (2016), who reported that while the first cycle of open data publication addresses transparency mainly for purposes of innovation, the second addresses issues of collaboration around government policies and toward greater accountability for policy performance. Ruijer et al. (2017) then explored specific contexts of open data usage to support the complexity of democratic processes, namely monitorial, deliberative and participatory democratic. On the other hand, making data available does not automatically yield up transparency (Bogdanović-Dinić et al., 2014; Matheus and Janssen, 2015; Zuiderwijk et al., 2012).

The first sub-dimension is focused on enabling greater openness of this process for stakeholders. However, it should be noted that openness always indicates a particular strategy defined by data producer (owner) in which the emphasis is on what the producer really wants to open. It results in different dominant categories of data published on open data portals, especially at regional and local levels of a country. As reported by Matheus and Janssen (2015), openness reveals details of what, how, and why politics took the decision, without revealing important parts of the political game inside government. Tkacz (2012) then considered the recent proliferation of openness as a political concept, because once an organisation, state or project is labelled open, it becomes difficult for closed politics to emerge from within. A model for assessing data openness with real-world application capabilities was introduced by Bogdanović-Dinić et al. (2014). Their model evaluates government's openness in the context of participation and collaboration features of open data portals and provides a link between transparency and engagement dimensions.

Secondly, there is an accountability sub-dimension. Its fulfilment requires the involvement of stakeholders to constitute a basis on which they may interact with each other. Their relationship is based on the control of data producers (owners) by other stakeholders. Matheus and Janssen (2015) emphasized the financial and operational aspects of this control. The main limitation is related to human capacities, skills, and knowledge of public actions. Zuiderwijk et al. (2012) reported that opening and using data encounters numerous impediments that can have both a socio and a technical nature. Therefore, it is crucial to support the engagement processes in terms of impacting on fostering accountability relationships between stakeholders. On the other hand, there is a lack of tools and a lack of trained personnel to properly work with these data. As a result, some of the most pressing challenges are associated with the tools and techniques used. Further, concerns over privacy and national security that may hinder the accountability efforts have received considerable attention of researchers (Chen et al., 2014; Klievink et al., 2017; Leuprecht et al., 2016). Janssen and van den Hoven (2015) argued that "*only by ensuring transparency the public knows what government is doing with BOLD and only by ensuring privacy citizens can be free from fear to express their opinions and vote.*"

Finally, the last sub-dimension is conditioned by legitimacy and ethics of public agencies and institutions' behaviour. This often arises when a stakeholder requests for disclosure of data (Matheus and Janssen, 2015; Reggi and Dawes, 2016). Thus, legitimacy and ethical factors require a focus of attention on the involvement of stakeholders in the process of designing, monitoring, and enforcing BOLD analytics initiatives and regulations. According to Dwivedi et al. (2017), ethics should be also considered in the context of tools as well as policies and guidelines are needed that are capable of ensuring the privacy and security of data. This advocates for the formation of online communities and engagement platforms in promoting compliance.

Engagement Dimension

An engagement dimension contains activities that should be realized by stakeholders and result in promoting data reuse and innovation, i.e. participation, collaboration, and cooperation. At first, there should an interest of stakeholders to participate in BOLD analytics. They have various, often conflicting, requirements as well as levels of expertise and knowledge. Therefore, it is appropriate to form groups according to such interests and enhance the process and outcomes of participation. However, participation must not be limited to only members of a particular group or organization. This effort may be viewed as "*a continuous process of empowerment and active involvement of all stakeholders in decision and policy-making processes with the use of ICT-powered data analytics platforms, tools, and services*." Most of researchers and practitioners recommend using various communication and delivery channels, participation platforms, tools or services to obtain public opinion and feedback on related initiatives, strategies, policies and particularly public services offered (Chatfield and Reddick, 2018; Janssen et al., 2012; Jetzek et al., 2013; Zuiderwijk et al., 2012; United Nations, 2016).

Further, the goal of the collaboration process is to "*enhance partnership between stakeholders by stimulating the development of new and innovative platforms, tools, and services that are shown to be effective in facilitating information, data and knowledge exchange.*" Collaboration is crucial to create value through innovation, new products and services, and decision-making models. It is also an important way to improve resource sharing, utilization and deployment, and develop collaborative networks. According to Hilbert (2016), collaboration pushes the implementation of open data forward. Cooperation then means "*providing these initiatives, strategies, policies, and resources across stakeholders.*" In particular, cooperation has to be supported across all public agencies and institutions, especially on regional and local levels, where it is important to coordinate activities and reduce duplication, overlap, and fragmentation among them (Attard et al., 2016; Máchová and Lněnička, 2017). More cooperation among citizens means more data available for BOLD analytics purposes, which should be used by public agencies and institutions to adapt their activities, services and decision-making to citizen demands (Klievink et al., 2017).

Failure to properly address these tasks could lead to an increase of costs to opening up data, inefficient use of computing resources, delays in data processing, worsen discoverability, accessibility and reusability of these data by stakeholders, etc. (Kučera et al., 2013; Máchová and Lněnička, 2017; Martin et al., 2017; Zuiderwijk et al., 2012). It is crucial to gain a better insight in who does what, where, with whom, and with which results, to enable monitoring of the BOLD analytics lifecycle. The issue of value co-creation with the usage of BOLD is an important application domain for this approach since building a network of stakeholders enables to share knowledge and skills to drive innovation, develop new products and services, and improve decision-making processes.

Legal Dimension

Legal dimension deals with legal conditions under which BOLD analytics may be used. Thus, public agencies and institutions must comply with the current legal framework, which contains the main regulations, principles, and requirements on privacy policy for data disclosure, licences and conditions of use, and protection of systems, i.e. data infrastructure including data portals (Kalampokis et al., 2013). The legal framework establishes the boundaries for activities involving in the engagement dimension. In this context, the most discussed issue is to ensure data privacy and security. The BOLD analytics lifecycle consists of several phases that incorporate various privacy and security requirements (Attard et al., 2016; Chen et al., 2014; Kučera et al., 2013). In the last step, only non-privacy-restricted and non-confidential data have to be disclosed to public. In this regard, Hardy and Maurushat (2017) argued that a preferable approach is to focus on achieving more reliable de-identification of government data than to relax existing privacy protections.

However, there is still a research gap on data anonymization techniques and in which phase or on which data should be applied. Reasons to exclude data depend on each institution, but can vary from security concerns to the need to protect confidential information (Dwivedi et al., 2017; Hardy and Maurushat, 2017). Although most countries have adopted privacy policies that govern the collection, storage, use and disclosure of data, their interpretation often differs across public agencies and institutions. Consequently, no one can guarantee that if these policies are implemented, the required data will be published on data portals. Open and linked data require licences and legal conditions under which these data are provided to be reused by stakeholders. There are a lot of open licences that allow different levels of reuse. Some countries create their own licenses focusing on their legal environment (Máchová and Lněnička, 2017). In this case, a license should be compatible with other open licenses. In addition, the BOLD analytics ecosystem has to deal with the General Data Protection Regulation (GDPR) compliance that requires coordinated approach to data management of personal data.

Finally, the protection of systems that operate the majority of a country's data (critical) infrastructure is a primary concern for governments. Consequently, policies require the protection of systems that contain sensitive information (Höchtl et al., 2016; Martin et al., 2013). It is especially important in the case of open data portals that can be used as backdoors to penetrate other information systems or databases. Another concerns related to security are distributed processing and storing of large amounts of data and using cloud computing resources in the public sector. Both these issues have received considerable critical attention in the literature. Nowadays, BOLD analytics is one of the strategies to deal with cybersecurity and critical infrastructure (Klievink et al., 2017; Leuprecht et al., 2016; Lněnička et al., 2017).

Technical Dimension

The technical dimension of BOLD analytics deals with establishing a new government enterprise architecture framework. Enterprise architecture is the collection of business processes, applications, technologies, and data that support the business strategies of an organization. Its goal is to create a unified ICT environment, i.e. promote alignment, standardization, reuse of existing ICT assets, and the sharing of common methods for software development across the organization (Minoli, 2008). A government architecture is a relatively young discipline in which concepts are slowly emerging. Often terms and concepts are used in various ways and there is no uniform agreement on these concepts (Janssen et al., 2013). Furthermore, while specific reasons for enterprise architecture adoption in the government may vary, however, according to Ojo et al. (2012) common reasons include: (1) enabling interoperability and providing technical and managerial standards for public agencies and institutions, (2) enabling resource sharing across them and reducing the cost of ICT and operations by identifying opportunities for reuse, and (3) enabling the development of shared processes and delivery of seamless services.

Therefore, it is important to integrate the relevant platforms, tools, and services for BOLD analytics and ICT standards in existing information systems and related processes as these are the key elements of the enterprise architecture. In addition, these elements have specific requirements on computing resources, which give public agencies and institutions an option to provide scalable, flexible, and cost-effective data infrastructure (Gonzalez-Zapata and Heeks, 2015; Hilbert, 2016; Vossen, 2014). It has to be robust, secure and able to store all these data as well as provide resources for open data portals when datasets are transferred across the networks and shared by the various systems and applications (Chen et al., 2014; Hilbert, 2016; Lněnička and

Komárková, 2015; Vossen, 2014). Interoperability of data infrastructure with other systems and applications has to be also ensured (Zuiderwijk et al., 2012). The intention of those requirements is to provide an improved data infrastructure within government in which data adhere to all quality standards (Gonzalez-Zapata and Heeks, 2015).

At the same time, this approach makes various demands on datasets, their metadata, formats, etc., especially in the case of linked data on the Web (Attard et al., 2016; Gonzalez-Zapata and Heeks, 2015; Heath and Bizer, 2011). From a technical perspective, these requirements must be compared with existing policies, so that the approach can be implemented at the different administrative levels. Shadbolt et al. (2012) consider how to bring open government data into the Web of Data. They reported that ease of access and better infrastructure are critical to realize this value. As a result, the big investments made by governments for the development of data infrastructures makes it necessary to evaluate them systematic (Charalabidis et al., 2014; Máchová and Lněnička, 2017).

A central point of data infrastructure accessible for public is represented by an open data portal. It can be defined as "*a website that serves as a catalogue of available datasets disclosed under open licenses and, at the same time, as a platform to engage stakeholders*." These portals are great resources for innovation and growth through collaborative value creation, particularly when dealing with linking these data (Reggi and Dawes, 2016). Some practitioners also recommended connecting national open data portal with regional or local data portals, i.e. it should act as an aggregator of these portals (Máchová and Lněnička, 2017). The portals are launched by an official government entity or a citizen initiative (Kalampokis et al., 2013; Veljković et al., 2014) and are required to be highly available services that provide reusable data, which are universally available and consumable (Millette and Hosein, 2016).

On the other hand, Ubaldi (2013) reported that many governments focus on the development of a national open data portal as if it were a higher priority than developing technical infrastructures to open up public data for others to use. Thus, another important challenges arising is the matter of data infrastructure for these data, i.e. networks, storage and servers. This issue is often solved using cloud computing technologies, because cloud computing and BOLD are complementary approaches to data storing, processing, analysing, accessing, and reporting (Lněnička and Komárková, 2015; Vossen, 2014). Regarding the design of these data portals, Ruijer et al. (2017) found that each type of democratic process requires a different approach aimed at ensuring effective use of open data. Finally, it should be also noted that there is an increasing interest in smart technologies and their applications in BOLD analytics (Edelenbos et al., 2018; Lněnička et al., 2018). In such a context, smart technologies are characterized by "*a transition to some extent intelligent electronic devices or systems enabling broad access to relevant information and knowledge that should help people to get informed and improve the decision-making process.*" Edelenbos et al. (2018) discussed these efforts in the context of the complexity of smart data cities and concluded that these data are key drivers of physical, economic and social change.

Social Dimension

The information society is characterized by the creation, storage, processing, distribution, and integration of data into economic, political, cultural, and social activities of stakeholders with the use of ICT. Since the main driver of this kind of society is the data exchange through interactions and interconnections between and among networks of stakeholders, BOLD analytics represents a major opportunity for the development of the data-driven economy.

At the same time, the potential to create value from these data is closely related to a level of capacities, knowledge, and skills of stakeholders (Chen et al., 2014; Kapoor et al., 2015; Weerakkody et al., 2017). Even if opening up data lowers the barriers for stakeholders to discover, access, and reuse them, the value creation often depends on a specific set of technical skills. Obviously, each of the BOLD analytics lifecycle phases requires different levels of them, but they have to be clearly defined at the beginning of the whole process. More precisely, these requirements should be already fulfilled in the legal dimension and related initiatives, strategies, and policies. For example, open data portals should provide high quality metadata about datasets as well as offer basic analytics and visualization tools to facilitate data reuse (Attard et al., 2016; Máchová and Lněnička, 2017). Finally, Dwivedi et al. (2017) suggested to focus on these elements of human resource factors in the BOLD analytics ecosystem: leadership, management competency, knowledge, capacity building, and symmetry of information.

On the other hand, these disproportions and inequalities may contribute to the widening gap of the digital divide. Hilbert (2016) stated that the result is a new kind of digital divide: a divide in data-based knowledge to inform intelligent decision-making. Thus, it is important to ensure that these data will comply with open licenses and other ICT standards enabling their efficient discoverability, accessibility, and reusability for stakeholders. There are a number of factors identified in the literature that influence the existence of the digital divide. They are mostly related to differences in income, education, literacy, broadband coverage, availability of technology, etc. (United Nations, 2016). These issues may be overcome by the support of the engagement dimension. Some recommendations are presented in Hilbert (2016).

Economic Dimension

Although differences in opinion still exist, there appears to be some agreement that the most important aim is to create value from these data. The economic potential lies in reducing costs of gaining information and providing the private sector with a tool to generate employment and revenue (Barry and Bannister, 2014; Chen et al., 2014; Janssen et al., 2012; Jetzek et al., 2013; Miller and Mork, 2013; Zuiderwijk et al., 2014; Vossen, 2014). In addition, the use of big data analytics enables public sector agencies and institutions to co-create public values with citizens (Chatfield and Reddick, 2018).

The creation of value is the core purpose and central process of the economic dimension. This can be thought of as a sort of data analytics lifecycle where platforms, tools or involved stakeholders in succeeding processing steps add value to these data (Attatd et al., 2016; Jetzek et al., 2014; Lněnička and Komárková, 2015). As mentioned above, BOLD analytics consists of different phases in which the activities are performed. Their contribution and importance in value creation varies. Some researchers emphasize data processing and analysis as the key phase (Debattista et al., 2015; Günther et al., 2017; Miller and Mork, 2013), while the others consider data publication, sharing, and reusing as the most important value generating mechanisms (Attatd et al., 2016; Charalabidis et al., 2014; Jetzek et al., 2014). However, both these phases are complementary since the output of each phase is provided as a feedback to the other. Moreover, if raw data are not processed and analysed thoroughly, then the quality of published data will be insufficient for further reuse and value creation.

As noted by Jetzek et al. (2013), there are two viewpoints concerning the value potential of these data: the first of them focuses on the economic value (e.g. contribution to the development of new e-services and mobile applications), while the second one focuses on the social value (e.g. contribution to improvements of the quality of lives of individuals or society as a whole through better government policies). Both these viewpoints involve innovation processes in which stakeholders create new products and services as well as form new decision-making models for further activities performed on these data. In a follow up paper, Jetzek et al. (2014) argued that open data sharing, and reuse can empower new ways of generating value in the sharing society. Their findings support the interrelation between economic and social dimensions.

4.2 BOLD Analytics Lifecycle

The increasing data requirements in the BOLD era bring various challenges on each lifecycle phase. However, although there are a number of data lifecycles, none of them are tailored to the specific needs of the BOLD analytics concept. Besides that, a number of activities are omitted, particularly the activities of data processing and analysis or data visualization. Therefore, all the phases and activities in the BOLD analytics lifecycle have to be covered in order to provide a standard methodology that can be followed by involved stakeholders. This description is not meant to be extensive, as each phase can also require other activities to be completed before the beginning of the following phase. In most cases, the main reason is data quality or security restrictions.

The BOLD analytics lifecycle addresses the key phases and activities that leverage the work with BOLD. Therefore, the following formal definition was given by Lněnička et al. (2016): "*The BOLD analytics lifecycle is governed by the principles of open government movement that is being fostered by increasing amounts of available data to be acquired and extracted, managed and prepared, stored and archived, processed and analysed, visualized and used, and published, shared, and reused."* These are the steps that transform raw data from data silos to freely available and linkable datasets for stakeholders, which will reuse them to create value. By describing these steps, the unique benefits, challenges and risks of BOLD may be related to specific data-related activities. Many of these activities are likely to already be present within most organizations. Table 5 summarizes them, including examples of literature sources. The lifecycle is visualized in Figure 2, providing a link to the dimensions of the BOLD analytics ecosystem.

Table 5: BOLD analytics lifecycle phases and activities

Phase	Activities	Example literature sources
Acquisition and extraction	Generate (create)	Attard et al. (2015); Charalabidis et al. (2016); Chen et al. (2014); Lněnička and Komárková (2015); Zuiderwijk et al. (2014)
	Select and filter	Di Martino et al. (2014); Tsai et al. (2015)
	Collect (gather)	Attard et al. (2015); Chen et al. (2014); Di Martino et al. (2014); Janssen and Kuk (2016); Kaisler et al. (2013); Miller and Mork (2013); Tsai et al. (2015); Zuiderwijk et al. (2014)
	Extract and load	Cárdenas et al. (2013); Lněnička and Komárková (2015); Miller and Mork (2013)
Management and preparation	Clean and prepare	Attard et al. (2015); Charalabidis et al. (2016); Chen et al. (2014); Di Martino et al. (2014); Kalampokis et al. (2013); Miller and Mork (2013); Zuiderwijk et al. (2014)
	Pre-process and validate	Charalabidis et al. (2016); Chen et al. (2014); Kaisler et al. (2013); Miller and Mork (2013); Tsai et al. (2015)
	Transform (format)	Charalabidis et al. (2016); Lněnička and Komárková (2015); Miller and Mork (2013); Tsai et al. (2015)
Storing and archiving	Store and secure	Attard et al. (2016); Cárdenas et al. (2013); Charalabidis et al. (2016); Chen et al. (2014); Di Martino et al. (2014); Janssen and Kuk (2016); Kaisler et al. (2013); Zuiderwijk et al. (2014)
	Transport and transfer	Chen et al. (2014); Janssen and Kuk (2016); Kaisler et al. (2013); Lněnička and Komárková (2015)
	Search and find	Lněnička and Komárková (2015); Zuiderwijk et al. (2014)
	Archive and curate	Attard et al. (2015); Charalabidis et al. (2016); Di Martino et al. (2014); Lněnička and Komárková (2015)

Processing and analysis	Analyse and integrate	Charalabidis et al. (2016); Chen et al. (2014); Di Martino et al. (2014); Kaisler et al. (2013); Lněnička and Komárková (2015); Miller and Mork (2013); Tsai et al. (2015)
	Model and simulate	Di Martino et al. (2014); Janssen and Kuk (2016); Miller and Mork (2013)
	Predict and optimize	Di Martino et al. (2014); Janssen and Kuk (2016); Lněnička and Komárková (2015)
Visualization and use	Visualize and interact	Charalabidis et al. (2016); Di Martino et al. (2014); Lněnička and Komárková (2015); Kalampokis et al. (2013); Miller and Mork (2013); Zuiderwijk et al. (2014)
	Evaluate and interpret	Attard et al. (2016); Di Martino et al. (2014); Tsai et al. (2015); Zuiderwijk et al. (2014)
	Export and report	Attard et al. (2015); Janssen and Kuk (2016); Lněnička and Komárková (2015)
Publication, sharing, and reuse	Search and retrieve	Attard et al. (2016); Charalabidis et al. (2016); Lněnička and Komárková (2015); Máchová and Lněnička (2017); Zuiderwijk et al. (2014)
	Define standards (licenses)	Attard et al. (2015); Charalabidis et al. (2016); Janssen and Kuk (2016); Janssen et al. (2012); Lněnička and Komárková (2015); Zuiderwijk et al. (2014)
	Interlink and connect	Attard et al. (2015); Charalabidis et al. (2016); Debattista et al. (2015); Kalampokis et al. (2013); Zuiderwijk et al. (2012); Zuiderwijk et al. (2014)
	Mashups (combine)	Attard et al. (2015); Debattista et al. (2015); Janssen and Kuk (2016); Janssen et al. (2012); Lněnička and Komárková (2015); Zuiderwijk et al. (2014)
	Collaborate and discuss	Charalabidis et al. (2016); Máchová and Lněnička (2017); Susha et al., 2015; Zuiderwijk et al. (2012); Zuiderwijk et al. (2014)
	Request and suggest	Máchová and Lněnička (2017); Susha et al., 2015; Zuiderwijk et al. (2012); Zuiderwijk et al. (2014)
	Share and spread	Attard et al. (2016); Máchová and Lněnička (2017); Zuiderwijk et al. (2012); Zuiderwijk et al. (2014)

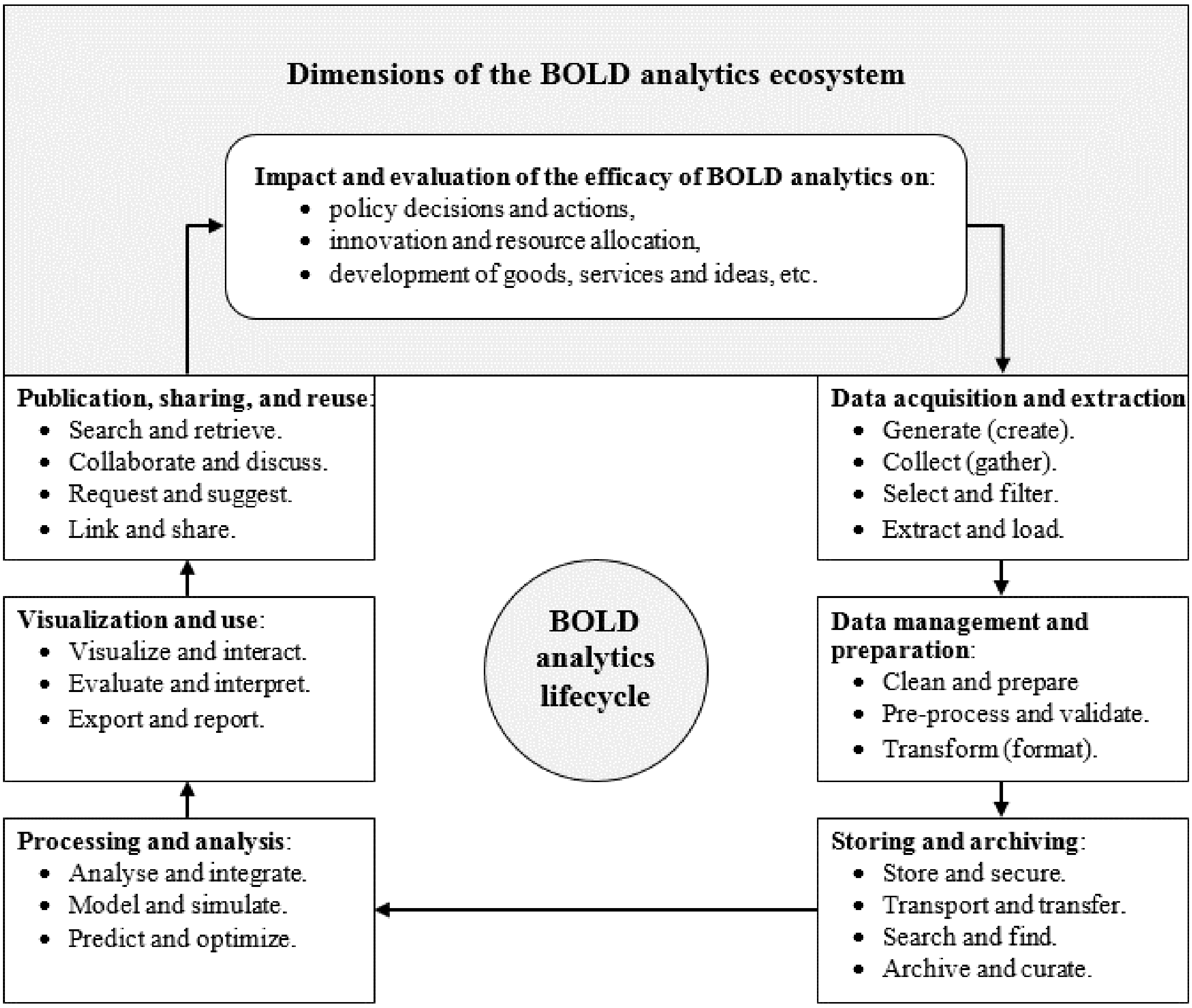

Figure 2: The lifecycle in the context of the BOLD analytics ecosystem's dimensions

4.3 Stakeholders in the BOLD Analytics Ecosystem

The primary effort of the BOLD analytics concept is to make more data from public sector agencies and institutions readily accessible to diverse stakeholders with a view of reusing them and foster participation, collaboration, and cooperation. According to Kapoor et al. (2015), this should contribute to the public policy-making space. Janssen et al. (2012) then argued that opening up data should result in open government in which the government acts as an open system and interacts with its environment, i.e. involved stakeholders. In this context, a stakeholder is a person, group or organisation that has an interest in, or is potentially impacted by, the operations of the public sector and its agencies or institutions. However, as reported by Peng et al. (2016), effective work with data and data products requires an integrated and coordinated team effort in multiple domains. Therefore, multiple stakeholders from different disciplines and diverse practices have to be brought together to examine the relationships between types of data (Janssen and Kuk, 2016).

In this regard, it is essential to identify the key stakeholders in the BOLD analytics ecosystem. In contrast to previous research, their necessity for this ecosystem is exemplified together with role(s) they have in the ecosystem, what their interests and requirements are and in which phases of the BOLD analytics lifecycle they are active. The stakeholder approach provides a framework that enables to define relationships between them. This should result in improving the participation, collaboration, and cooperation processes as well as capabilities, knowledge, and skills of stakeholders to create value from these data. This approach is widely used in the literature to define roles and formalize responsibilities that will facilitate the process of data analytics including curating and communicating quality information to stakeholders (Gonzalez-Zapata and Heeks, 2015; Peng et al., 2016)

Public agencies and institutions, citizens, non-profit organisations, and businesses are the key stakeholders who can interact in this ecosystem. However, as stated by Attard et al. (2016), since the efforts of the latter stakeholders remain largely uncoordinated, their roles, motivations, levels of expertise, and priorities differ. According to Zuiderwijk et al. (2014), open data ecosystems are expected to bring many advantages, such as stimulating citizen participation and innovation. In this ecosystem, ICT enables the discussion of open data and stimulates the participation of citizens in governmental processes of decision and policy-making. The best functionalities of this ecosystem can be selected and utilized by open data providers and users. Heimstädt et al. (2014) conceptualized open data ecosystems by analysing the major stakeholders in the UK. They identified a set of structural business ecosystem properties: circular flow of resources, sustainability, demand that encourages supply, and dependence developing between suppliers, intermediaries, and users. Ubaldi (2013) identified three types of interacting open data ecosystems, namely 1) an ecosystem of data producers, 2) an ecosystem of infomediaries as intermediate consumers of data and 3) an ecosystem of open data users.

The literature provides various views on this issue. Van den Broek et al. (2011) assigned five internal stakeholder roles to the various steps of the lifecycle: top management, information manager, legal advisor, community manager, and data owner. Martin et al. (2013) distinguished citizens, public administration, politics and industry as the main stakeholder groups and additionally media and science. Thereupon user types are derived, namely: producer and publisher, user and consumer, and prosumer. Attard et al. (2016) identified six roles in which stakeholders can participate to create value: data producer, data enhancer, data publisher, service creator, facilitator, and data consumer. According to Dawes et al. (2016), the ecosystem encompasses the involvement of three stakeholder groups: (1) government leaders and organizations responsible for open data programs, (2) direct open data users who comprise expert data analysts and members of the communities who develop applications with these data, and (3) the beneficiaries of data reuse, comprising both individuals and organizations who adopt, buy, and use the products and services based on these data.

Another point of view is distinguishing relevant stakeholders into two groups: primary (politicians, public officials, public sector practitioners, international organizations) and secondary (ICT providers, academics, funding donors, civil society activists), see Gonzalez-Zapata and Heeks (2015). Martin et al. (2017) then emphasized the importance of the stimulator who can set goals for the ecosystem, not only on what it can produce, but also on its form and functioning. Finally, Charalabidis et al. (2014) reported that there is no more a clear distinction between providers and consumers of these data. As a result, they emphasized the importance of data prosumers, i.e. users who both produce and consume data.

Based on the findings gathered from the literature review, the following shortcomings may be observed: (1) stakeholders responsible for policy-making processes and mechanisms of data disclosure are not taken into account, (2) control and intermediary elements (roles) in the ecosystem are often missing, (3) external roles are preferred over internal roles in the public sector, (4) the importance of data infrastructure for value creation is underestimated as well as the availability of platforms, tools, and services to work with these data. Therefore, a new classification of stakeholders and their roles in the BOLD analytics ecosystem is developed and described by Figure 3.

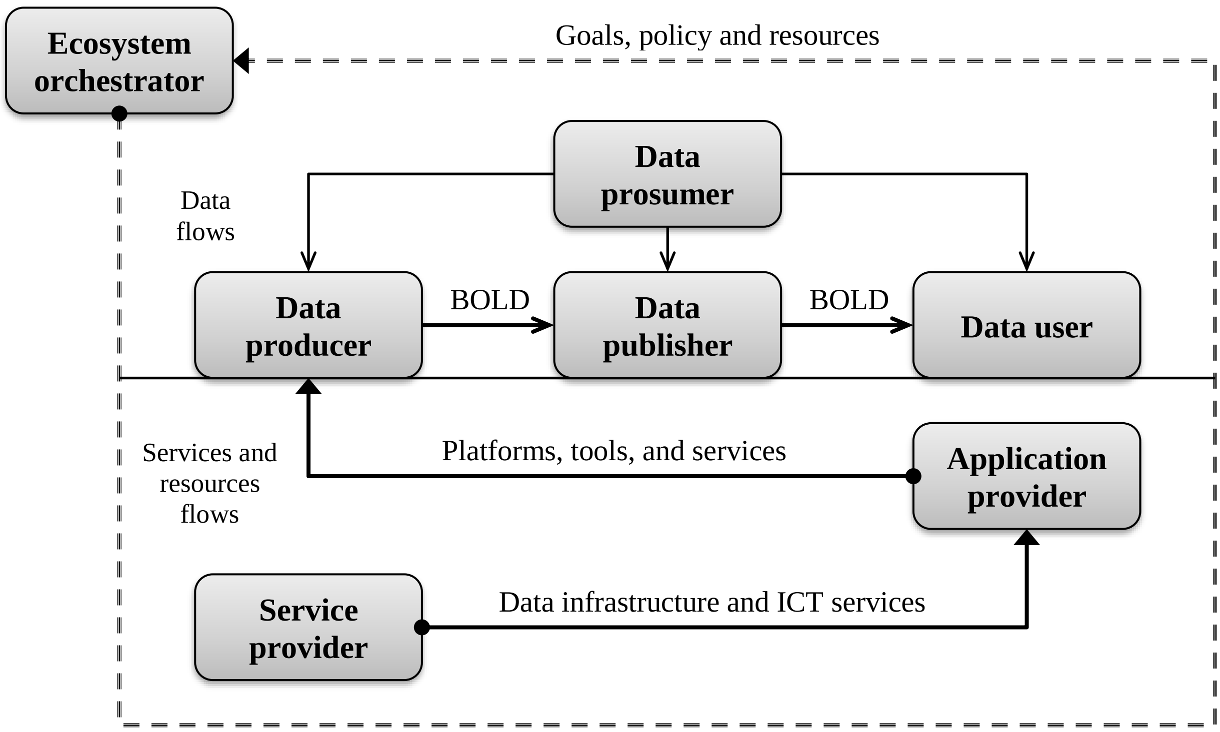

Figure 3: The main stakeholders in the BOLD analytics ecosystem

The first stakeholder is characterized as an ecosystem orchestrator who may be represented by parliament, president, politicians, ministries, bureaucrats, courts, etc. Their responsibility lies in defining an ecosystem by regulation and enforcement of required activities and operational guidelines in the context of organization's goals, policy and resources. An orchestrator establishes the structure and legal boundaries of the ecosystem for resources and services delivery and data flows between other stakeholders. Since the ecosystem is open, the orchestrator has to deal with pressures from internal as well as external stakeholders to achieve efficient and sustainable use of ecosystem services. Moreover, various ecosystems supporting BOLD analytics can be created by regional or local authorities.

The second stakeholder is a service provider who may be represented by businesses or public agencies and institutions. The main activity is establishing and maintaining a computing framework in which certain applications will be executed. Most often it is about providing data infrastructure and ICT services, guaranteeing their availability, scalability, reliability, and security. Service provider is the first type of intermediary who ensures the transport of services between stakeholders at different levels.

The third stakeholder is an application provider (enhancer) who may be represented by public agencies and institutions or universities and research centres. This stakeholder is responsible for the BOLD analytics lifecycle management and the execution of related activities using platforms, tools, and services. The main activities include utilizing them to transform raw data to be disclosed. This stakeholder is important because of the BOLD analytics complexity and the required level of technical skills to work with these data.

The fourth stakeholder in the BOLD analytics ecosystem is a data producer (owner). This role can be considered as one of the most important roles within the ecosystem. The traditional role of governments as data producers (owners) is facing a competition with universities and research centres or non-governmental organizations. Production of these data involves their generation and maintenance in the course of daily operations of an organization. The main responsibility is to ensure adequate data quality and compliance with standards, including the protection of private and confidential data, for further activities. Data producers make their data readily available to the data publisher.

Then, there is a data publisher (intermediary) who may be represented by public agencies and institutions or non-governmental organizations. This stakeholder may be the same as data provider. However, findings suggest that there should be a control element that is responsible for data disclosure channels, feedback from stakeholders, and terms of data licensing. When new datasets are published, publishers have to make these datasets known to potential data users to promote the creation of new products and services. They usually advertise on various communication and delivery channels, facilitate the discoverability of these data, and support engagement processes. Data publisher is the second type of intermediary who guides the data distribution between providers and consumers.

The sixth stakeholder is a data user (consumer) who may be represented by individual citizens, communities, businesses, public agencies and institutions, universities and research centres or non-governmental organizations. Their main activities involve providing feedback about data quality or reusability for public agencies and institutions and creating value from these data. Data users are in the centre of this ecosystem.

Finally, the last stakeholder is a data prosumer, i.e. user who produces as well as consumes data. It is represented by citizens, businesses or non-governmental organizations that combine various datasets, including their own data, and publish their updates on the Web. Some open data portals allow registered users to publish these data on the portal.

Each of these stakeholders also participates in different phases of the BOLD analytics lifecycle. Together, they ensure the continuity of each phase by means of activities and required outcomes. This can be seen from Table 6. The ecosystem orchestrator influences all the phases throughout decision and policy-making processes. The service provider then provides data infrastructure and ICT services for all the phases. The application provider performs related activities using platforms, tools, and services. However, this stakeholder does not participate in data publication, sharing, and reuse. The data producer creates data and stores them, or transfers them from an information system into a database. The data publisher processes these data and prepares them for publication in compliance with standards and licences. He also participates in the phase of visualization and use in which creates reports and overviews to be published together with datasets. These should help users to get a proper summarization and choose the most suitable dataset. The data user then reuses these data and creates value. The data prosumer acts as both data producer and user.

Table 6: The BOLD analytics lifecycle phases in which each stakeholder participates within the ecosystem

	Acquisition and extraction	Management and preparation	Storage and archiving	Processing and analysis	Visualization and use	Publication, sharing, and reuse
Ecosystem orchestrator	YES	YES	YES	YES	YES	YES
Service provider	YES	YES	YES	YES	YES	YES
Application provider	YES	YES	YES	YES	YES	NO
Data producer	YES	NO	YES	NO	NO	NO

Data publisher	NO	YES	YES	YES	YES	YES
Data user	NO	NO	NO	NO	NO	YES
Data prosumer	YES	NO	YES	NO	YES	YES

4.4 BOLD Analytics Framework Proposal

A conceptual framework for BOLD analytics was developed to show how the identified elements and their relationships are represented in the ecosystem (see Figure 4). The most important requisites of the BOLD analytics ecosystem are robust technical and technological (hardware) infrastructure, generic (software) services and platforms, and human capacities and skills. Because, as stated by Hilbert (2016), they can be unequally distributed, leading to a development divide. These horizontal layers are then employed to analyse different aspects of the BOLD analytics lifecycle (vertical layers of Figure 4). The phases and related activities are identified above. Data storage and archiving phase supports all the other phases. It also means that these phases and their activities have to be addressed for all the horizontal layers. For example, computing resources meeting a set of requisites for BOLD processing and analysis, including availability, scalability, reliability, and security, are needed to ensure that the value from these data will be unlocked. Furthermore, technical capacities and skills are needed to use generic services and platforms, which are able to provide techniques for visualization and use of these data.

From a system-theoretic perspective, public policy choices can broadly be categorized into two groups (Hilbert, 2016). The first one is dealing with positive feedback on the ecosystem in the means of incentives that foster specific dynamics. It is represented by the mechanisms of transparency and engagement dimensions providing processes for stakeholders to exploit the data for innovative purposes. The second one then considers the effects of negative feedback, such as legal and technical regulations that curb particular dynamics. These aspects such as security, GDPR, quality standards, and other requirements may restrict access to data or disclose data with low quality (Dwivedi et al., 2017; Hilbert, 2016; Lněnička et al., 2016; Sayogo et al., 2014). These are the diagonal layers of Figure 4. Beyond the framework's boundaries are its context and external environment (externalities). This is important as the whole concept of an ecosystem implies that it is open to its externalities and other ecosystems existing in the public sector. It should be noted that the identified elements are not independent on the locational, cultural, and social environments or other ecosystems in which they are used but are interdependent with them in meeting the needs of the public sector. We acknowledge that perceptions and fit with an organization are crucial to the successful development and implementation of the BOLD analytics ecosystem.

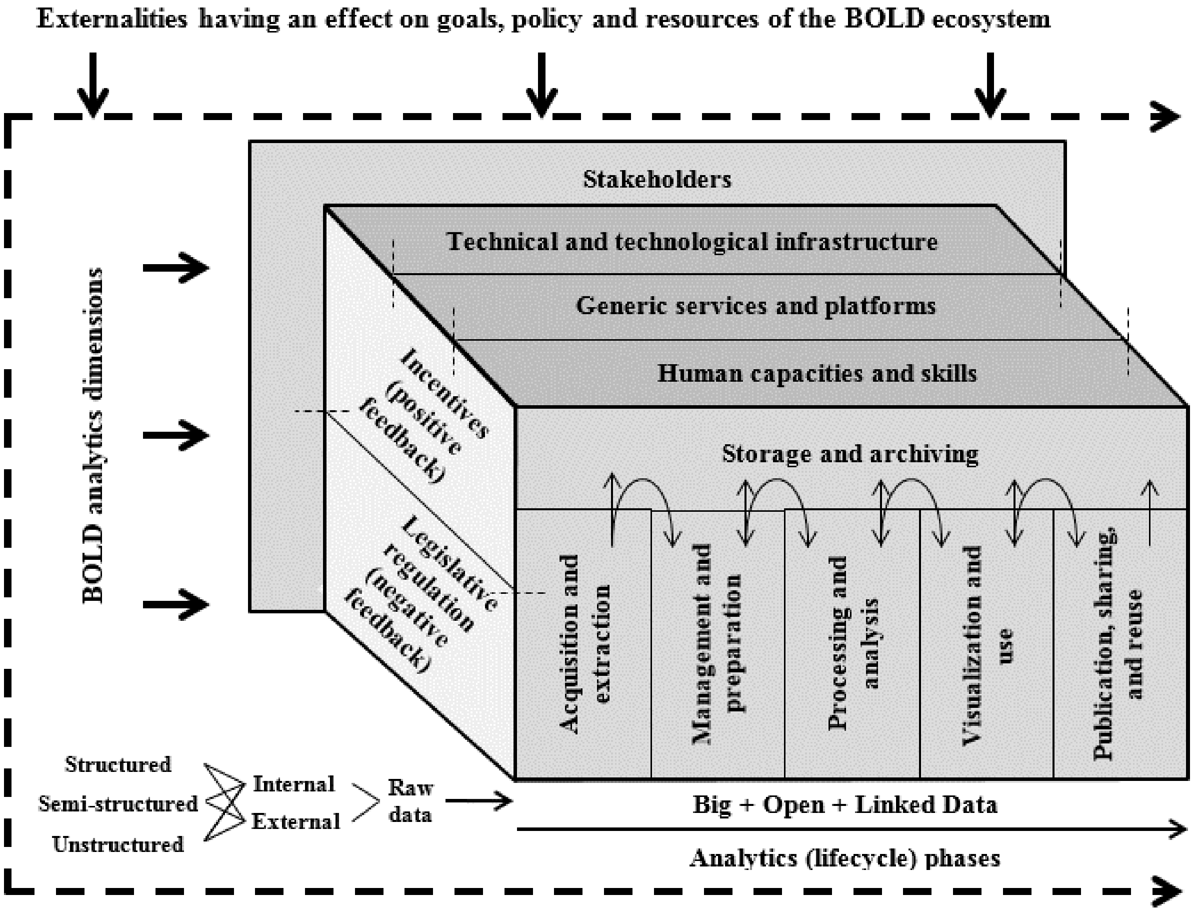

Figure 4: The conceptual framework of the BOLD analytics ecosystem

The result is a three-dimensional conceptual framework of the BOLD analytics ecosystem, which was firstly employed by Hilbert (2016) and later extended by Lněnička et al. (2016). It consists of different circumstances (horizontal) and interventions (diagonal) intersecting and affecting different applications of the BOLD analytics lifecycle phases and activities (vertical). The lifecycle is cyclical, reflecting the continuous nature of BOLD where value and innovation through BOLD feed into each other (Dwivedi et al., 2017; Jetzek et al., 2014). The internal boundaries of this ecosystem are shaped by the BOLD analytics dimensions in which the value is created. Different stakeholders are active in the BOLD analytics lifecycle phases. As the main asset of the ecosystem, data may be structured, semi-structured or unstructured in nature. Internal data come from within the public sector and are produced, collected or held by public sector agencies and institutions. External data come from various sources outside of the public sector, such as international organizations, businesses, researchers, etc. Raw data are unprocessed computer data that have not been subjected to processing or any manipulation. The resulted datasets should be available as linked data, which make them easier to discover, access, and reuse by stakeholders. It is worth mentioning that the order of the phases and activities might be changed based on particular needs of organizations.

The BOLD analytics ecosystem is represented by a cube corresponding to the layers. Each of them represents an important level of requirements within BOLD analytics. To achieve this, scalable, flexible, and efficient robust technical and technological infrastructure, generic services and platforms, and particularly coordination of human capacities and skills are required to address the challenges within the phases and activities of the BOLD analytics lifecycle carried out by different stakeholders. Thus, by formalizing BOLD analytics in terms of its essential elements, the ecosystem clarifies this concept, making it more understandable, communicable, and practical for both practitioners and policymakers. In addition, it should foster the development of these ecosystems in the public sector.

4.5 BOLD Analytics Framework Validation and Benefits

To validate the proposed conceptual framework, a comparison with existing frameworks was made. Generally, it can be stated that the main differences lie in their narrower focus describing a particular data analytics perspective rather than the whole ecosystem in which the coexistence of these perspectives is discussed and described. Sayogo et al. (2014) conceptualized a framework consisting of only two elements helping to promote openness: data manipulation and engagement capability. Only the perspective of open data was also explored by Heimstädt et al. (2014), but they argued that there is a need to generate interoperable data allowing for extensive cross-case analysis. In other words, data should be linked to easily gather more data for big data analytics.

Only big data analytics was explored by Hilbert (2016). The conceptual framework of this author thus did not reflect the requirements of open and linked data. On the other hand, since the framework of Hilbert considered the concept of public policy choice, this layer can be extended to reinforce the relationships within BOLD analytics and related dimensions. As stated above, this change is required to bridge the different views on the BOLD analytics requirements. Although Dwivedi et al. (2017) emphasized the importance of technical infrastructure in their framework encouraging innovation through BOLD, they omitted most of the BOLD analytics lifecycle phases. Only access to data (storing, retrieving or using data), together with security (protecting data from destructive forces and from the unwanted actions of unauthorised users) and privacy (confidentiality of sensitive information) were taken into account. This shortcoming is overcome by the new conceptual framework of the BOLD analytics ecosystem.

The usefulness of this framework for the public sector can be demonstrated by extending the focus of existing frameworks to include the elements of the BOLD analytics ecosystem. Höchtl et al. (2016) dealt with the applicability of a big data-enabled model for the large-scale interpretation of public opinion to be used in policy-making. Although their research is focused on the public sector, they presented only the point of view within public agencies and institutions. In this regard, our conceptual framework provides a list of stakeholders that should be involved in the decision-making process.

Furthermore, the benefits expecting from these data cannot be achieved without discussing the requirements of each step required to collect, manage, process, visualize, and reuse them. Our conceptual framework offers a detailed description of the phases and activities of the BOLD analytics lifecycle. Those were also not completely covered by other authors such as Janssen and Kuk (2016), Kaisler et al. (2013), Matheus and Janssen (2015) and Miller and Mork (2013). In most cases, these authors described only the activities related to data processing and analysis, but did not explore the opportunities provided by data publication, sharing, and reuse. Although it can be argued that they did not explicitly keep the focus on transparency or engagement efforts, dissemination and exploitation of the results are also very important activities, especially in the project management context.

The framework can also provide insights that can be useful in improving a linked open government data analytics approach proposed by Kalampokis et al. (2013), which was based on a case study that is related to the general elections of the United Kingdom. Especially, it can be helpful in identifying the relationships between involved stakeholders and the phases and activities of the BOLD analytics lifecycle. The relations can be also mapped to the ecosystem of stakeholders introduced by Martin et al. (2017), since the authors argued that there is a need to "*identify and develop the interdependencies, to increase the density and the strength of the relationships* *among the elements of the ecosystem*." In this regard, our conceptual framework can impulse new dynamics among the elements of the ecosystem by providing a more complex view than single perspectives of data analytics. Only one or two perspectives were applied by Peng et al. (2016), Reggi and Dawes (2016), Shadbolt et al. (2012); Susha et al. (2015), Van den Broek et al. (2011), Veljković et al. (2014), Weerakkody et al. (2017) or Zuiderwijk et al. (2014).

Incorporating some of the essential elements of the BOLD analytics ecosystem presented in our conceptual framework may also improve other existing frameworks, such as those discussed in the previous paragraphs. Since the interconnection between different data analytics perspectives cannot be overlooked, it is important to redesign existing frameworks with this in mind. Thus, our conceptual framework consists of an overview of related dimensions discussing the benefits and risks associated with BOLD analytics, which should be taken into account by governments and public agencies and institutions. Other layers are represented by the most important requisites and public policy choices affecting the BOLD analytics ecosystem together with the key phases and activities of the BOLD analytics lifecycle and stakeholders ensuring that all these elements will be managed by people with domain and technology expertise.

5 Discussion and Limitations

It is necessary to understand the essential elements of the BOLD analytics ecosystem if involved stakeholders wish to work with these data. In this regard, this article attempts to fill this gap by conducting the comprehensive literature review assessing the existing evidence base for this topic. At first, the boundaries of this ecosystem are represented by six dimensions. Each of these dimensions is described in terms of its sub-dimensions and related benefits, risks, and requirements to success. However, other dimensions may be considered to be important for the ecosystem.

For example, Vossen (2014) considered the organizational dimension in which an adoption strategy for these data is discussed. These requirements should be a part of the legal dimension since governments are responsible for policy execution. Some authors also emphasized the importance of one dimension over the others in the ecosystem. For example, Gonzalez-Zapata and Heeks (2015) argued that bureaucratic and political perspectives are more dominant than technological and economic perspectives. Since the authors found out that the explanation is related to the capacities and interests of key stakeholders, it may be suggested that the reason is the focus on either requirements of big data (technical dimensions) or open data (transparency dimension). Reflecting the potential of e-government to support the implementation of sustainable development goals (United Nations, 2016), an ecological /environmental dimension may be incorporated into the framework. BOLD can support these goals by improving the granularity, comparability and accuracy of statistics on respective issues such as poverty, climate action and reducing inequality. In our framework, these efforts are encompassed within social, economic, and technical dimensions.

Although there are a number of data lifecycles, none of them are tailored to the specific needs of the BOLD analytics concept, i.e. the incorporation of requirements for big/open/linked data into a single lifecycle is missing. Therefore, the phases and activates of the BOLD analytics lifecycle were clearly identified and described. The exact number of phases that might be involved in BOLD analytics, the goals of these phases and their outcomes have gained the attention of many authors and thus the main limitation may come from these various points of view. However, the description is not meant to be extensive, as each phase can also require other activities to be completed before the beginning of the following phase. The importance of the stakeholders and their interests are also discussed very often. In this article, their necessity for the ecosystem is exemplified together with role(s) they have in the ecosystem, what their interests and requirements are, and in which phases of the BOLD analytics lifecycle they are active.

Further, fully exploring the potential of these data requires the development of stakeholders' capacities, skills and ability to find, interpret, and reuse them. These skills should include an understanding of the legal and technical requirements in relation to BOLD including their interpretation and limits, together with more detailed explanation as well as where supporting resources can be found. In addition, governments also need to promote the use of these data within public agencies and institutions at national, regional, and local level. This will need investment in platforms, tools and necessary training to be able to implement BOLD analytics. It should also mean that governments have to be leading and proactive stakeholders for supporting the reuse of these data, including advanced analytics services to improve internal decision-making and to help create new products and services.

According to Kaisler et al. (2013), management will, perhaps, be the most difficult problem to address with BOLD analytics. In addition, from a data-management perspective, Dwivedi et al. (2017) addressed these challenges: finding and dealing with large datasets, integrating datasets that were not intended to be integrated, or building usable data management interfaces for stakeholders of various levels of expertise. They argued that future research is required to uncover the effective techniques and data models to handle the integration of data. On the other hand, Hardy and Maurushat (2017) are cautious to research on collecting vast amounts of unconnected and disparate data for BOLD analytics since the use of predictive analytics may lead to controversial overreach by the state. They also argued that the greater the amount of data collected and analysed, the greater the risk that individuals may be re-identified from that data through cross-referencing. Similarly, it should be taken into account that access to these data can raise questions regarding their security, especially in relation to sensitive data and GDPR, and can also raise concerns about the trust of data, so leading to higher risk for using and implementing them further (Dwivedi et al., 2017).

The limitations of the methodology are related to the literature selection process, i.e. English papers only, only online papers, etc., and the fact that the framework and conclusions derived come purely from the extant literature, i.e. secondary research only, and have not been validated in any way. However, since this article is only the first step toward filling the gap in our understanding of the positive effects of BOLD analytics, we will conduct an international Delphi study to identify and achieve consensus on priorities for the framework and essential elements. We are also going to validate our conclusions in real life conditions, i.e. investigating the readiness of public agencies and institutions to adopt BOLD analytics. In addition, the framework will be aligned with other types of evidence representing the externalities of the ecosystem in the public sector that may hinder its use.

The conceptual framework of the BOLD analytics ecosystem then comprises all these essential elements. However, the question whether these are all the elements is still open to discussion. There are various ecosystems in the literature that can be compared to the new BOLD analytics ecosystem. In contrast to these ecosystems, this review article aimed to provide the theoretical background necessary to understand the unique benefits, challenges, and risks of BOLD analytics.

It is also evident that not all of public data have the same potential to support citizen engagement and not all of these data have the same relevance for participation, collaboration, and cooperation. Further research should be focused on different thematic data categories on open data portals and data stakeholders need for their analyses. Some of portals already offer statistics about datasets (Máchová and Lněnička, 2017). In the context of the engagement dimension, Foulonneau et al. (2014) reviewed the current approaches to encouraging the creation of services based on open data. They especially emphasized the advertising of datasets and applications on virtual community channels, the harmonization of metadata vocabularies, training and support of involved stakeholders by creating APIs on these portals, visualizing datasets in maps, and providing a large set of documentation and tools to assist them in data reuse.

The article does not deal with particular technical issues, at least not directly. Some comments on it can be found in previous work of authors, e.g. Lněnička and Komárková (2015). Google's MapReduce (or its open source implementation – Apache Hadoop) has been according to many authors successfully used for big data processing within different application fields in practice, e.g. analysis of large genomic data sets by Hadoop (O'Driscoll et al., 2013), description of the distribution of the traffic volume, subscribers, and requests among service providers in cellular data networks (Jun et al., 2013) or power grid time series data analysis (Bach et al., 2013). Various implementations of MapReduce were described by Goyal and Bharti (2015). Apache Spark (another engine for big data processing) was used to predict bank customer's behaviours (Etaiwi et al., 2017). Starfish is "*a self-tuning system for big data analytics*"; it was introduces by Herodotou et al. (2011) to automatically provide a good performance. EpiC system was designed to allow handling Big Data's variety (Jiang et al., 2016).

Related to the BOLD analytics lifecycle, another field that requires further work in the use of BOLD is a classification of platforms, tools, and services, as well as policies and guidelines that are capable of ensuring the privacy and security of these data (Lněnička and Komárková, 2015). Therefore, more research is needed to identify and classify the platforms, tools, and services that could be deployed in each phase of the BOLD analytics lifecycle for the public sector purposes. Moreover, future work should address the issue of necessary hard skills that are required to perform BOLD analytics.

6 Conclusions

The ecosystem approach applied in this article demonstrated that it can be widely used to identify essential elements and their relationships addressing requirements of BOLD analytics in the public sector. Thus, our first contribution is the introduction of the essential elements of the BOLD analytics ecosystem. A second contribution lies in the fact that we explored, through empirical examples from the comprehensive literature review, how these elements act together in the ecosystem. Finally, a third contribution is the conceptual framework considering dimensions, lifecycle phases, and stakeholders, which was compared to existing frameworks and its usefulness for the public sector was discussed.

Since the BOLD analytics ecosystem involves the interaction of complex domains, its essential elements are presented in this article. The internal boundaries of this ecosystem are represented by the BOLD analytics dimensions that include transparency, engagement, legal, technical, social, and economic aspects of BOLD. More precisely, these boundaries are defined by dimensions that foster transparency of public agencies and institutions, enable and support engagement of related stakeholders, establish legal environment, maintain technical and technological infrastructure, assess social effects, and recognize economic impacts of BOLD analytics.

The second step in describing BOLD analytics is to differentiate activities in their lifecycle phases. The stakeholders in the BOLD analytics ecosystem have their roles and they are responsible for concrete phases. The conceptual framework of the BOLD analytics ecosystem then provides a comprehensive view on the essential elements and their interrelationships. With the use of this framework, related stakeholders can develop and implement BOLD services to ensure that the public agencies and institutions fulfil their tasks with quality, efficiency, and transparency.

The key benefit created by BOLD analytics is the increased speed of data distribution allowing policymakers and other stakeholders to reuse these data and take informed decisions. It may be concluded that BOLD analytics has the potential to lead to new scientific and research insights, create economic development opportunities, and generate new products and services that can contribute to the increase of government transparency. However, such actions will involve balancing the questions of access, privacy, security, data management, archiving, etc.

Finally, this article is significant because it offers useful insights into the essential elements of the BOLD analytics ecosystem. This article also examines the theoretical background of each dimension of BOLD analytics that can offer deeper insights into the benefits of using these data. The contribution to the practice of data analytics is shown through the identification of the most important phases. Moreover, this article fills an important gap in the extant literature on big/open/linked data. In this regard, it complements the stream of research that uses theoretical models to emphasize the importance of these data. As for future work, the validation of the proposed framework will be conducted to address whether the elements are comprehensive, explanatory, and accurate. For this purpose, an international Delphi study is going to be conducted.

References

Attard, J., Orlandi, F., & Auer, S. (2016). Data driven governments: Creating value through open government data. In A. Hameurlain et al. (Eds.), *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVII. Lecture Notes in Computer Science, 9860.* (pp. 84–110). Berlin Heidelberg: Springer. https://dx.doi.org/10.1007/978-3-662-53416-8_6.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399–418. https://dx.doi.org/10.1016/j.giq.2015.07.006.

Bach, F., Hueseyin, C., Maass, H., & Kuelinapfel, U. (2013). Power grid time series data analysis with Pig on a Hadoop cluster compared to multi core systems. *Proceeding of the 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing* (pp. 208–212). IEEE. https://dx.doi.org/10.1109/PDP.2013.37.

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, 19(1,2), 129–152. https://dx.doi.org/10.3233/IP-140327.

Bertot, J. C., & Choi, H. (2013). Big data and e-Government: Issues, policies, and recommendations. *Proceedings of the 14th Annual International Conference on Digital Government Research* (pp. 1–10). New York: ACM. https://dx.doi.org/10.1145/2479724.2479730.

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, 19(1, 2), 5–16. https://dx.doi.org/10.3233/IP-140328.

Bogdanović-Dinić, S., Veljković, N., & Stoimenov, L. (2014). How open are public government data? An assessment of seven open data portals. In M.P. Rodríguez-Bolívar (Ed.), *Measuring E-government Efficiency, Public Administration and Information Technology, 5.* (pp. 25–44). New York: Springer. https://dx.doi.org/10.1007/978-1-4614-9982-4_3.

Cárdenas, A. A., Manadhara, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy*, 11(6), 74–76. https://dx.doi.org/10.1109/MSP.2013.138.

Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2016). A taxonomy of open government data research areas and topics. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 41–63. https://dx.doi.org/10.1080/10919392.2015.1124720

Charalabidis, Y., Loukis, E., & Alexopoulos, C. (2014). Evaluating second generation open government data infrastructures using value models. *Proceedings of the 47th Hawaii International Conference on System Sciences, HICSS 2014* (pp. 2114–2126). IEEE. https://dx.doi.org/10.1109/HICSS.2014.267.

Chatfield, A. T., & Reddick, C. G. (2018). Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services. *Government Information Quarterly*, 35(2), 336–347. https://dx.doi.org/10.1016/j.giq.2017.11.002.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. https://dx.doi.org/10.1007/s11036-013-0489-0.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15–27. https://dx.doi.org/10.1016/j.giq.2016.01.003.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings*, 1644(1), 97–104. https://dx.doi.org/10.1063/1.4907823.

Debattista, J., Lange, C., Scerri, S., & Auer, S. (2015). Linked 'big' data: Towards a manifold increase in big data value and veracity. *Proceedings of 2015 IEEE/ACM 2nd International Symposium on Big Data Computing* (pp. 92–98). IEEE. https://dx.doi.org/10.1109/BDC.2015.34.

Di Martino, B., Aversa, R., Cretella, G., Esposito, A., & Kołodziej, J. (2014). Big data (lost) in the cloud. *International Journal of Big Data Intelligence*, 1(1-2), 3–17. https://dx.doi.org/10.1504/IJBDI.2014.063840.

Dwivedi, Y. K., Janssen, M., Slade, E. L., Rana, N. P., Weerakkody, V., Millard, J., ... & Snijders, D. (2017). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modelling. *Information Systems Frontiers*, 19(2), 197–212. https://dx.doi.org/10.1007/s10796-016-9675-5.

Edelenbos, J., Hirzalla, F., van Zoonen, L., van Dalen, J., Bouma, G., Slob, A., & Woestenburg, A. (2018). Governing the complexity of smart data cities: Setting a research agenda. In M.P. Rodríguez-Bolívar (Ed.), *Smart Technologies for Smart Governments, Public Administration and Information Technology, 24*. (pp. 35–54). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-58577-2_3.

Etaiwi, W., Biltawi, M., & Naymat, G. (2017). Evaluation of classification algorithms for banking customer's behavior under Apache Spark data processing system. *Procedia Computer Science*, 113, 559–564. https://dx.doi.org/10.1016/j.procs.2017.08.280.

Foulonneau, M., Turki, S., Vidou, G., & Martin, S. (2014). Open data in service design. *Electronic Journal of e-Government*, 12(2), 99–107.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. https://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007.

Geiger, C. P., & von Lucke, J. (2012). Open government and (linked) (open) (government) (data). *JeDEM-eJournal of eDemocracy and Open Government*, 4(2), 265–278.

Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32(4), 441–452. https://dx.doi.org/10.1016/j.giq.2015.09.001.

Goyal A., & Bharti (2015). Study on emerging implementations of MapReduce. *International Conference on Computing, Communication & Automation* (pp. 16–21). IEEE. https://dx.doi.org/10.1109/CCAA.2015.7148364.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems,* 26(3), 191–209. https://dx.doi.org/10.1016/j.jsis.2017.07.003.

Hardy, K., & Maurushat, A. (2017). Opening up government data for big data analysis and public benefit. *Computer Law & Security Review*, 33(1), 30–37. https://dx.doi.org/10.1016/j.clsr.2016.11.003.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the Web into a global data space. *Synthesis lectures on the semantic web: Theory and technology*, 1(1), 1–136.

Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1), 135–174. https://dx.doi.org/10.1111/dpr.12142.

Heimstädt, M., Saunderson, F., & Heath, T. (2014). Conceptualizing open data ecosystems: A timeline analysis of open data development in the UK. *CeDEM14: Proceedings of the International Conference for E-Democracy and Open Government 2014* (pp. 245–255). Krems: Edition Donau-Universität Krems.

Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B., & Babu, S. (2011). Starfish: A self-tuning system for big data analytics. *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research* (pp. 261–272), Asilomar: CIDR.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 147–169. https://dx.doi.org/10.1080/10919392.2015.1125187.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258–268. https://dx.doi.org/10.1080/10580530.2012.716740.

Janssen, M., Flak, L. S., & Sæbø, Ø. (2013). Government architecture: Concepts, use and impact. In M. A. Wimmer, M. Janssen, & H. J. Scholl (Eds.), *Proceedings of the 12th IFIP WG 8.5 International Conference, EGOV 2013* (pp. 135–147). Berlin Heidelberg: Springer. https://dx.doi.org/10.1007/978-3-642-40358-3_12.

Janssen, M. & Kuk, G. (2016). Big and Open Linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*. 26(1-2), 3–13. https://dx.doi.org/10.1080/10919392.2015.1124005.

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?. *Government Information Quarterly*, 32(4), 363–368. https://dx.doi.org/10.1016/j.giq.2015.11.007.

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). Generating value from open government data. *Proceedings of 34th International Conference on Information Systems, ICIS 2013* (pp. 1–20). Berkeley: Bepress.

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2014). Generating sustainable value from open data in a sharing society. In B. Bergvall-Kåreborn & P. Nielsen (Eds.), *Creating Value for All Through IT, IFIP Advances in Information and Communication Technology, 429*. (pp. 62–82). Berlin Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-662-43459-8_5.

Jiang, D., Wu, S., Chen, G., Chin Ooi, B., Tan, K.-L., & Xu, J. (2016). epiC: an extensible and scalable system for processing Big Data. *The VLDB Journal*, 25(1), 3–26. https://dx.doi.org/10.1007/s00778-015-0393-2.

Jun, L., Tingting, L., Gang, C., Hua, Y., & Zhenming, L. (2013). Mining and modelling the dynamic patterns of service providers in cellular data network based on big data analysis. *China Communications*, 10(12), 25–36. https://dx.doi.org/10.1109/CC.2013.6723876.

Kapoor, K., Weerakkody, V., & Sivarajah, U. (2015). Open data platforms and their usability: Proposing a framework for evaluating citizen intentions. In M. Janssen et al. (Eds.), *Open and Big Data Management and Innovation* (pp. 261–271). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-25013-7_21.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. *Proceedings of 2013 46th Hawaii International Conference on System Sciences, HICSS 2013* (pp. 995–1004). IEEE. https://dx.doi.org/10.1109/HICSS.2013.645.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Linked open government data analytics. In M. A. Wimmer, M. Janssen, & H. J. Scholl (Eds.), *Proceedings of the 12th IFIP WG 8.5 International Conference, EGOV 2013* (pp. 99–110). Berlin Heidelberg: Springer. https://dx.doi.org/10.1007/978-3-642-40358-3_9.

Khan, S., Liu, X., Shakil, K. A., & Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing & Management*, 53(4), 923–944. https://dx.doi.org/10.1016/j.ipm.2017.03.006.

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, 19(2), 267–283. https://dx.doi.org/10.1007/s10796-016-9686-2.

Kucera, J., & Chlapek, D. (2014). Benefits and risks of open government data. *Journal of Systems Integration*, 5(1), 30–41.

Kučera, J., Chlapek, D., & Nečaský, M. (2013). Open government data catalogs: Current approaches and quality perspective. In A. Kő, C. Leitner, H. Leitold, & A. Prosser (Eds.), *Technology-Enabled Innovation for Democracy, Government and Governance, Lecture Notes in Computer Science, 8061*. (pp. 152–166). Berlin Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-40160-2_13.

Lake, P., & Drake, R. (2014). *Information Systems Management in the Big Data Era*. Cham: Springer International Publishing.

Leuprecht, C., Skillicorn, D. B., & Tait, V. E. (2016). Beyond the Castle Model of cyber-risk and cyber-security. *Government Information Quarterly*, 33(2), 250–257. https://dx.doi.org/10.1016/j.giq.2016.01.012.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline*, 9(1), 181–212. https://dx.doi.org/10.28945/479.

Lněnička, M., Čapek, J., Komárková, J., Máchová, R., & Čermáková, I. (2017). A solution to combat cybersecurity threats involving big data analytics in the Hadoop ecosystem. *Proceedings of the 30th International Business-Information-Management-Association Conference* (pp. 1804–1812). Norristown: IBIMA.

Lněnička, M., & Komárková, J. (2015). The impact of cloud computing and open (big) data on the enterprise architecture framework. *Proceedings of the 26th International Business-Information-Management-Association Conference* (pp. 1679–1683). Norristown: IBIMA.

Lněnička, M., Komárková, J., Máchová, R., & Pásler, M. (2016). Formalizing the requirements of big open linked data analytics in the public sector. *Proceedings of the 28th International Business-Information-Management-Association Conference* (pp. 4453–4457). Norristown: IBIMA.

Lněnička, M., Máchová, R., Komárková, J., & Pásler, M. (2018). Government enterprise architecture for big and open linked data analytics in a smart city ecosystem. In V. Uskov, R. Howlett, & L. Jain (Eds.), *Smart Education and e-Learning 2017. SEEL 2017. Smart Innovation, Systems and Technologies, 75*. (pp. 475–485). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-59451-4_47.

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12(1), 21–41. https://dx.doi.org/10.4067/S0718-18762017000100003.

Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Open data: Barriers, risks and opportunities. In E. Ferrari, & W. Casternovo (Eds.), *Proceedings of the 13th European Conference on eGovernment, ECEG 2013* (pp. 301–309). Reading: Academic Conferences and Publishing International Limited.

Martin, S., Turki, S., & Renault, S. (2017). Open data ecosystems: Introducing the stimulator function. In A. Kő, & E. Francesconi (Eds.), *International Conference on Electronic Government and the Information Systems Perspective, EGOVIS 2017* (pp. 49–63). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-64248-2_5.

Matheus, R., & Janssen, M. (2015). Transparency dimensions of big and open linked data: Transparency as being synonymous with accountability and openness. In M. Janssen et al. (Eds.), *Open and Big Data Management and Innovation* (pp. 236–246). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-25013-7_19.

Miller, H. G., & Mork, P. (2013). From data to decisions: A value chain for big data. *IT Professional*, 15(1), 57–59. https://dx.doi.org/10.1109/MITP.2013.11.

Millette, C., & Hosein, P. (2016). A consumer focused open data platform. *Proceedings of the 2016 3rd MEC International Conference on Big Data and Smart City, ICBDSC* (pp. 1–6). IEEE. https://dx.doi.org/10.1109/ICBDSC.2016.7460350.

Minoli, D. (2008). *Enterprise Architecture A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology*. Boca Raton: Auerbach Publications.

O'Driscoll, A., Daugelaite, J. & Sleator, R. D. (2013). 'Big data', Hadoop and cloud computing in genomics. *Journal of Biomedical Informatics*, 46(5), 774–781. https://dx.doi.org/10.1016/j.jbi.2013.07.001.

Ojo, A., Janowski, T., & Estevez, E. (2012). Improving government enterprise architecture practice – Maturity factor analysis. *Proceedings of the 45th Hawaii International Conference on System Sciences, HICSS 2012* (pp. 4260–4269). IEEE. https://dx.doi.org/10.1109/HICSS.2012.14.

Peng, G., Ritchey, N., Casey, K., Kearns, E. J., Privette, J., Saunders, D., ... & Ansari, S. (2016). Scientific stewardship in the open data and big data era: Roles and responsibilities of stewards and other major product stakeholders. *D-Lib Magazine*, 22(5/6). https://dx.doi.org/10.1045/may2016-peng.

Ouyang Q., Wu, C., & Huang, L. (2018). Methodologies, principles and prospects of applying big data in safety science research. S*afety Science*, 101, 60–71. https://dx.doi.org/10.1016/j.ssci.2017.08.012.

Reggi, L., & Dawes, S. (2016). Open government data ecosystems: Linking transparency for innovation with transparency for participation and accountability. In H. J. Scholl et al. (Eds.), *Proceedings of the 15th IFIP WG 8.5 International Conference, EGOV 2016* (pp. 74–86). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-44421-5_6.

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, 34(1), 45–52. https://dx.doi.org/10.1016/j.giq.2017.01.001.

Saxena, S. (2017). BOLD (Big and Open Linked Data): what's next?. *Library Hi Tech News*, 34(5), 10–13. https://dx.doi.org/10.1108/LHTN-04-2017-0020.

Sayogo, D. S., Pardo, T. A., & Cook, M. (2014). A framework for benchmarking open government data efforts. *Proceedings of the 47th Hawaii International Conference on System Sciences, HICSS 2014* (pp. 1896–1905). IEEE. https://dx.doi.org/10.1109/HICSS.2014.240.

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, 27(3), 16–24. https://dx.doi.org/10.1109/MIS.2012.23.

Susha, I., Grönlund, Å., & Janssen, M. (2015). Organizational measures to stimulate user engagement with open data. *Transforming Government: People, Process and Policy*, 9(2), 181–206. https://dx.doi.org/10.1108/TG-05-2014-0016.

Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering*, 115, 319–330. https://dx.doi.org/10.1016/j.cie.2017.11.017.

Tkacz, N. (2012). From open source to open government: A critique of open politics. *ephemera*, 12(4), 386–405.

Tsai, C. W., Lai, C. F., Chao, H. C. & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*, 2(1), 1–21. https://dx.doi.org/10.1186/s40537-015-0030-3.

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives, *Working Papers on Public Governance*, 22, OECD Publishing.

United Nations. (2016). *United Nations e-Government Survey 2016: E-Government in Support of Sustainable Development*. New York: UN Publishing Section.

Van den Broek, T., van Veenstra, A. F., & Folmer, E. (2011). Walking the extra byte: A lifecycle model for linked open data. In E. Folmer, M. Reuvers, & W. Quak (Eds.), *Linked Open Data – Pilot Linked Open Data Nederland* (pp. 95–111). Amersfoort.

Van der Waal, S., Węcel, K., Ermilov, I., Janev, V., Milošević, U., & Wainwright, M. (2014). Lifting open data portals to the data Web. In S. Auer, V. Bryl, & S. Tramp (Eds.), *Linked Open Data – Creating Knowledge Out of Interlinked Data* (pp. 175–195). Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-09846-3_9.

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), 278–290. https://dx.doi.org/10.1016/j.giq.2013.10.011.

Vossen, G. (2014). Big data as the new enabler in business and other intelligence. *Vietnam Journal of Computer Science*, 1(1), 3–14. https://dx.doi.org/10.1007/s40595-013-0001-6.

Wang, J., Crawl, D., Purawat, S., Nguyen, M., & Altintas, I. (2015). Big data provenance: Challenges, state of the art and opportunities. *Proceedings of the 2015 IEEE International Conference on Big Data* (pp. 2509–2516). IEEE. https://dx.doi.org/10.1109/BigData.2015.7364047.

Weerakkody, V., Kapoor, K., Balta, M. E., Irani, Z., & Dwivedi, Y. K. (2017). Factors influencing user acceptance of public sector big open data. *Production Planning & Control*, 28(11-12), 891–905. https://dx.doi.org/10.1080/09537287.2017.1336802.

World Wide Web Foundation. (2015). *Open Data Barometer Global Report: Third Edition*. Washington DC: Web Foundation.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1,2), 17–33. https://dx.doi.org/10.3233/IP-140329.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R. S., & Sheikh Alibaks, R. (2012). Socio-technical impediments of open data. *Electronic Journal of e-Government*, 10(2), 156–172.