# Government Big Data Ecosystem: Definitions, Types of Data, Actors, and Roles and the Impact in Public Administrations

SYED IFTIKHAR HUSSAIN SHAH, School of Science & Technology, International Hellenic University, Thessaloniki, Greece
ASST. PROF. VASSILIOS PERISTERAS, European Commission, DG Informatics, Brussels, Belgium, and International Hellenic University, Thessaloniki, Greece
IOANNIS MAGNISALIS, School of Science & Technology, International Hellenic University, Thessaloniki, Greece

8

The public sector, private firms, business community, and civil society are generating data that are high in volume, veracity, and velocity and come from a diversity of sources. This type of data is today known as big data. Public administrations pursue big data as "new oil" and implement data-centric policies to collect, generate, process, share, exploit, and protect data for promoting good governance, transparency, innovative digital services, and citizens' engagement in public policy. All of the above constitute the Government Big Data Ecosystem (GBDE). Despite the great interest in this ecosystem, there is a lack of clear definitions, the various important types of government data remain vague, the different actors and their roles are not well defined, while the impact in key public administration sectors is not yet deeply understood and assessed. Such research and literature gaps impose a crucial obstacle for a better understanding of the prospects and nascent issues in exploiting GBDE. With this study, we aim to start filling the above-mentioned gaps by organizing our findings from an extended Systematic Literature Review into a framework to organise and address the above-mentioned challenges. Our goal is to contribute in this fast-evolving area by bringing some clarity and establishing common understanding around key elements of the emerging GBDE.

CCS Concepts: • **Information systems → Data management systems**; **Information systems applications**; • **Human-centered computing →** *Collaborative and social computing*; • **General and reference →** Surveys and overviews;

Additional Key Words and Phrases: Big data, big data actors and roles, data and information, data-driven government, government big data ecosystem

Syed Iftikhar Hussain Shah is a Researcher, and Ph.D. Candidate in Electronic Government/Information Systems Management at International Hellenic University, Greece.
The European Union-funded Project: Digital Europe for All (DE4A), Horizon 2020—the Framework Programme for Research and Innovation (2014-2020), H2020-SC6-GOVERNANCE-2018-2019-2020 (Governance for the Future), Grant Agreement 870635, and the European Union-funded project Co-Inform, Horizon 2020 the Framework Programme for Research and Innovation (2014-2020) H2020-SC6CO-CREATION-2016-2017 (Co-creation for Growth and Inclusion), Grant Agreement 770302.
Authors' addresses: S. I. H. Shah and I. Magnisalis, School of Science & Technology, International Hellenic University, Thessaloniki, Greece; emails: {i.shah, i.magnisalis}@ihu.edu.gr; Asst. Prof. V. Peristeras, European Commission, DG Informatics, Brussels, Belgium, and International Hellenic University, Thessaloniki, Greece; email: v.peristeras@ihu.edu.gr.

1936-1955/2021/05-ART8 $15.00
https://doi.org/10.1145/3425709

**ACM Reference format:**
Syed Iftikhar Hussain Shah, Asst. Prof. Vassilios Peristeras, and Ioannis Magnisalis. 2021. Government Big Data Ecosystem: Definitions, Types of Data, Actors, and Roles and the Impact in Public Administrations. *J. Data and Information Quality* 13, 2, Article 8 (May 2021), 25 pages.
https://doi.org/10.1145/3425709

## 1 INTRODUCTION

Big data is sweeping across numerous areas of public and private organizations. Becker [9] emphasises that "big data is the oil of the 21st century," as the capacity to exploit big data has become a critical success factor for Public Administrations (PAs), private firms, and civil society.

Recently, data-driven public and private organizations define data policies and draft data strategies aligned with the vision, mission, and goals of their organization. They also try to secure expertise and skills in data-related fields and put in place data technologies in areas covering data generation and collection, storage, access, flow, sharing, publishing, management, analysis, re(use), protection, privacy, and preservation [2, 27, 28].

Data-driven transformation for large organizations is an intensive, resource-consuming, and long-term endeavor that affects technology, human resources, organizations, and cultures [73]. Data champions extensively use big data and data analytics in every aspect of their business, including sales, marketing, supply chain, manufacturing, Human Resource management, and Research & Development (R&D) [27, 28, 73]. In data-driven organizations, data are considered a key strategic asset [41]. Big data usually refers to data characterized by the four Vs., i.e., large volume, veracity, variety, and velocity. The ability to process big data creates opportunities for organizations to achieve competitive benefits in the current digitized marketplace [81]. This ability needs cost-effective and distinctive modern big data tools and techniques to be put in place [65, 76]. The exploitation of big data represents a paradigm shift in tactics to understand and study the world [4].

In the public sector, the adoption of big data technologies and approaches provides opportunities, including efficient public service delivery, data-driven decision making for policymakers, progress in the digital economy, creation of new jobs, promotion of civic participation in the definition, and improvement of public policies [4, 58, 92].

A Big Data Ecosystem (BDE) is a complex set of numerous interconnected components related to big data, models, and organisational structures and roles covering the entire data lifecycle [23]. It comprises different components, including data infrastructure, data analytics, data models, as well as organisational and cultural elements [22, 23, 56].

In our research, we have identified four research gaps in the study of the GBDE: (a) there is no well-established definition of the GBDEs, (b) no holistic work on the classification of types of government data, (c) no harmonization of data actors and their roles, and (d) no clarity on the impact on the various PA sectors. In this study, we focus on addressing the above-mentioned research gaps by organizing specific findings under a framework. Such aspects are the fundamental elements of the GBDE and are identified as follows: (a) a definition of GBDEs, (b) a classification of government big data types, (c) a classification of government big data actors and their roles, and (d) the use and potential impact of big data in the core business processes of various key PAs areas. To create our framework, we conducted a Systematic Literature Review (SLR) [21, 46, 47]. We provide details about our approach in the forthcoming sections of this study.

The remainder of the research article is structured as follows. In Section 2, we illustrate the background of our research work. Section 3 explains our research methodology. Section 4 represents the results of the literature review. Last, in Section 5, we present research implications and the limitations of this study, conclude our study, and present proposed future work.

## 2 BACKGROUND AND SCOPE OF THIS WORK

In this section, we explain the background of this research, and we present the research gaps that have attracted our interest.

### 2.1 Overview of the GBDE Field

The word "data" originates from the Latin word "datum" [73]. Data are a discrete, limitless entity that have an unstructured and unprocessed shape. Organizations further process such kinds of data, as per their needs, to illustrate relevant objects, events, concepts, or facts [73].

Big data is data in high volume, veracity, velocity, and variety. Big data needs economical, advanced ways of information processing to be used for generating insight and supporting decision making [39, 65]. Presently, organizations, including public organizations, are flooded with a massive quantity of big data generated with high speed [73], through "smart" data sources. These big data sources include people, the Internet, smart mobile handsets, online social networks (Twitter, Facebook, LinkedIn, Instagram), the Internet of Things (IoT), Global Positioning Systems, and so on [10, 26, 51, 60, 86, 95]. There are over 2.77 billion social network users and about 26.66 billion IoT connected devices and sensors worldwide [95, 92]. The generated data relate to all public sectors, e.g., health, education, agriculture, transportation, and social welfare [86, 95, 92].

The "ecosystem lens" for studying this phenomenon is particularly useful to understand interdependencies among collaborators in exchange networks [35, 80]. The BDE reveals a complex, connected ecosystem of high-capacity networks, data users and data applications, and services needed to store, process, and visualize data that are gathered from multiple data sources [14, 96]. The BDE stakeholders include public and private organizations, development partners, civil society, and users. This ecosystem can help PAs to make evidence-based decisions, ensure data interoperability and data privacy, prioritize needs and problems, encourage civic participation, and perhaps contribute to a better government [58, 92]. As such, the study of this BDE is a new field of growing importance.

### 2.2 Research Gaps

We identified the following research gaps while studying the GBDE.

*2.2.1 RG1: No well-established definition of Government Big Data Ecosystems.* We found about 22 research studies, which had endeavored to define the concept of a "data ecosystem." The authors of the research papers utilized the term "data ecosystem" without giving a coherent and consistent definition [6, 31, 97]. Some authors [22, 98] referred to generic ecosystems' definitions in their research work. For example, they refer to general ecosystems [6, 35, 74, 80], business ecosystems [52], information system ecosystems [13], software ecosystems [59], and digital ecosystems [33]. We try to draft a definition using the existing studies.

*2.2.2 RG2: No Extensive Work on the Classification of Types of big Government Data.* In the literature, researchers primarily discover the technological features of the BDE [23, 67]. There are a few studies, e.g., References [1, 15, 94, 69], about different types of big data, and each study proposes different typologies based on their particular research subject. We found literature discussing types of government big data in scattered form. For example, Reference [61] discusses personal data [23] and differentiates between structured and unstructured data, and Reference [60] focuses on open versus closed data. We did not find research proposing robust classifications and holistic typologies for big data [92].

*2.2.3 RG3: Confusing Literature on the Data Actors and Roles.* In the literature, we found numerous types of data actors and roles. However, many variations of actors and their roles appear due to using different actors' titles and contradictory roles by the researchers [38, 48, 61, 98]. The research was dispersed in various studies to identify data actors, distinct roles, communities, actors' relationships, and data actors' motivation. We did not find a proposal for an all-inclusive typology of data actors and roles.

*2.2.4 RG4: Lack of Research to Identify and Assess the Impact of Big Data in Different Public Administrations Sectors, i.e., Education, Health, Transport, Security, and so on.* It appears that big data has an impact in all PA sectors, including education, justice, budgeting, policymaking, economic, agriculture, safety and security, and transportation. We hardly found a study that specifically highlighted the potential impact in the above-mentioned sectors.

## 3 RESEARCH METHOD

In the preceding section, we identified four research gaps while reviewing the GBDE. In this study, we aim to mitigate these gaps and organize certain findings under a framework to include (a) a definition of GBDE, (b) a classification of government big data actors and their roles, (c) a classification of government big data types of big data, and (d) the impact of big data in key PAs sectors.

To address the research gaps, we conducted qualitative research about the GBDE by using the SLR. SLR is a research method to identify, evaluate, and interpret research work, literature produced by scholars, researchers, and practitioners [5]. Fink defined literature review as a systematic, specific, and reproducible approach to identify, evaluate, and synthesize the existing research work delivered by scholars, researchers, and practitioners [5]. We pursued this research methodology following guidelines from the literature [21, 46, 47]. We detail our approach in five steps. The SLR process or research review protocol's *first step* is centered on formulating the research questions. The *second step* mainly focuses on three sub-activities that include the selection of digital research libraries, the formulation of search strings, and literature search. The *third step* is mainly centered on the identification of the relevant studies and applying quality assessment on the studies. This is supported by defining and applying inclusion and exclusion criteria. The *fourth step aims* to analyze the studies, extract relevant information, perform verification of outcomes, and link the findings to research gaps. The *last step* explains the research results and organizes them in our proposed framework for the GBDE. Figure 1 summarises the steps. We briefly present them in the subsequent sections below.

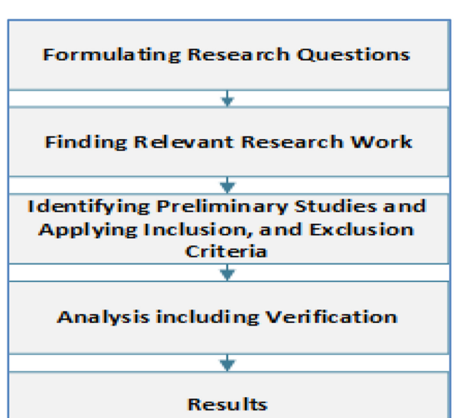


Fig. 1. Steps in the SLR Process.

### 3.1 Goal and Research Questions

This study goals to address the research gaps by organizing our findings and proposing a framework. The mitigation of these gaps may be considered as fundamental elements of the GBDE. To combine the fundamental elements of the GBDE within a comprehensive framework, we outlined the following Research Questions (RQs):

RQ1: What is a comprehensive definition of a BDE in the government context?
RQ2: What elements constitute a basic classification of big data types from the government perspectives?
RQ3: What are the data actors and their roles in government settings?
RQ4: What are the different areas of PA in which big data has potential impacts?

Comprehensive research work on the above areas is vital to establish a groundwork and common understanding of GBDE amongst scholars and practitioners. We barely find any research literature examining big government data as an ecosystem. Our proposed framework, based on the fundamental elements of GBDEs, has the clear benefit of building a common ground among the stakeholders of such ecosystems. Such stakeholders are, but not limited to, governments, citizens, businesses, scholars, and practitioners.

### 3.2 Finding Relevant Research Work

SLR applies a thorough approach to review the findings presented in prior published research. The focus of literature research is to carry out a systematic review that demands extensive coverage of current research performed on the research topic of interest within the stipulated period. Most of the researchers who attempted to work out a systematic review opinioned in a research survey that they did not stop off their literature search activity until they believed they had achieved their target [21]. This section elaborates about the above-mentioned SLR process second step that focuses on providing the following details about our selected digital research libraries, the procedure for the identification of search strings, and the searching process as well.

*Selection of digital research libraries:* To proceed with the SLR, the selection of digital research libraries is a challenging task. In this task, researchers usually decide about where to search and how to search for requiring research articles that contain relevant information for the research questions of the study [75]. We selected and investigated the following four digital research libraries for the literature search: ACM, IEEE Xplore, Science Direct, and Springer Link.

*Formulation of search string:* We formulated search strings to find research articles from the above-mentioned selected digital research libraries based on the following measures:

- We formulated search string related to the above-mentioned RQs,
- In the case of the critical aspect, like data actor, we find alternate words and synonyms for these keywords
- We use Boolean operators like ’OR’, ’AND’ to extend the search by adding other words and synonyms.
- In some cases, we also adjust/alter the search string.
- Each search string includes one or more than one keywords like “data actor” or “data ecosystem” player.

The above-mentioned measures are applied to formulate search strings for our research questions. For example, we describe the following as a search string.

Query: What is the classification of government big data actors and their roles?

The search string that is derived for the above-mentioned query is "data actor" OR "big data actor," OR "data player" OR "big data player" OR "actor in data ecosystem" OR "actors and roles" OR "roles in big data" OR "classification of data actors."

*Literature search:* We began the literature search to obtain relevant research papers in February 2019 and carried out this procedure until December 2019. All results from searches are mainly based on titles, keywords, and abstracts. We utilized the above-mentioned measures about the formulation of search strings for our literature search process. To get relevant literature about fundamental elements of GBDE, we performed a literature search activity in the following two stages.

Stage-I: We used the aforementioned four digital libraries to search strings and their variants, which are based on the following keywords:- "DATA ECOSYSTEM," "DATA ACTORS," "GOVERNMENT DATA ECOSYSTEM," "DATA ACTORS ROLES," "CLASSIFICATION OF DATA ACTORS," DATA PROVIDER," "DATA USERS," "DATA ACTOR," "BIG DATA ACTOR," "DATA PLAYER," "BIG DATA PLAYER," "DATA BUSINESS ENTITY," "DATA SUPPORT SERVICE PROVIDER," "Data ECOSYSTEM," 'Big Data TYPE," "DATA TYPES IN DATA ECOSYSTEM," "CLASSIFICATION OF BIG DATA TYPES," "POTENTIAL IMPACT OF DATA ECOSYSTEM," "IMPACT OF BIG DATA IN KEY AREAS OF PUBLIC ADMINISTRATION," DEFINITION OF DATA ECOSYSTEM," "POTENTIAL AREAS OF BIG DATA ECOSYSTEM," "USE AND POTENTIAL IMPACT OF BIG DATA IN GOVERNMENT," "DATA-DRIVEN GOVERNMENT, along with choices "exact phrase" and "matches all." We examined the outcomes of the above-mentioned first stage and matched the results with our crucial research sub-topics regarding GBDE. We observed that yet, we required additional relevant research papers, and then we decided to perform the following stage-II.

Stage-II: In this stage, we extended the search queries carried out in stage-I by adding "matches any" instead of options "exact phrase" and "matches all."

In total, we collected 904 research articles. We stored the literature search results in a spreadsheet where every row relates to a research paper. We captured various attributes and metadata per paper like paper ID, authors, title, authors, source, keywords, abstract, year of publication, unique viewer tag, searching date, associated search term, the aim of the study.

### 3.3 Identifying Preliminary Studies and Article Quality Assessment

The procedure to identify preliminary studies, apply inclusion and exclusion criteria, and perform a quality assessment is based on the following actions. We strictly followed our inclusion and exclusion criteria, given below, to assess the relevance of the studies with our research targets. We manually performed all the next steps to identify preliminary studies. Subsequently, we completed a thorough scrutiny process in the following three phases. We explain our inclusion and exclusion criteria, and then we describe the three phases of our process.

*Inclusion criteria*: To select only the relevant research articles, we included resources that satisfy one or more of the following criteria:

- discuss big data regarding the definition of the BDE
- focus big data regarding the definition of the BDE
- discuss types of big data regarding the classification of types of big data
- focus types of big data regarding the classification of types of big data
- discuss big data actors and their roles regarding the classification of big data actors and their roles
- focus big data actors and their roles regarding the classification of big data actors and their roles
- focus big data regarding potential impacts in PA

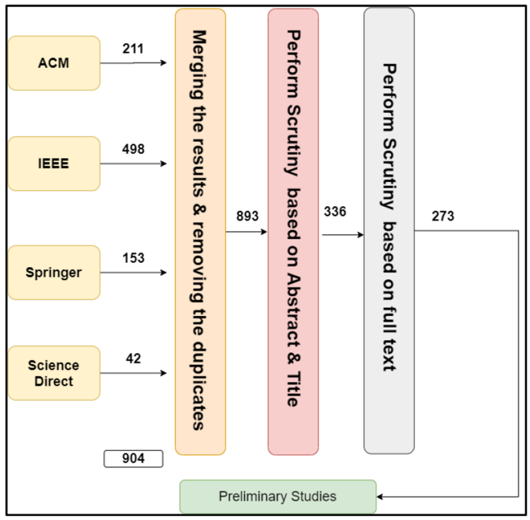


Fig. 2. Procedure for the identification of preliminary studies.

- discuss big data, government, data-driven government
- resources publication year range is not restricted and keeps it open
- depict applicable results outside of the study

*Exclusion Criteria*: We used the following exclusion criteria to filter out research articles that:

- are not written in English
- have no relevance to the central theme of the RQs
- have no primary focus or discussion on fundamental elements of GBDEs as per the aim of this study. For example, resources that are mainly focused on big data technologies that include Big Data Analytics, SQL-on-Hadoop systems, Machine Learning
- do not discuss on fundamental elements of GBDEs including, its potential impacts in PA

*Scrutiny process:* In this section, we explain the following major phases to scrutinize the research articles along with the quality assessment of the articles.

*Remove duplication:* We merged and placed the research articles that were found in the preceding literature search phase in a single shared folder and removed the research papers that are duplicate. From the above-mentioned literature search process, we collected a total of 904 research articles. In this stage, we removed duplications from our study dataset. It decreased the papers to a total of 893 research articles.

*Initial scrutiny based on Abstract and Title:* Initially, we examined the research articles based on abstract and title. If a research paper was not judged for inclusion or exclusion based on these traits, then it was added for the next step of review of this phase. In the next level of this phase, titles and abstracts were independently evaluated by two researchers. Each researcher noted the research articles that have some confusion to decide about the research article's inclusion or exclusion for the next scrutiny phase. Both researchers mostly found similar results, and there was minimal disagreement about inclusion and exclusion of papers between them. However, both researchers held meetings to resolve differences and to discuss disputed and marginal research articles.

In the completion of this phase of scrutiny, we reduced the papers to a total of 336 of 893 research articles.

*Scrutiny based on the full text:* In this phase, the research team reviewed the full text of articles that are already agreed in the above-mentioned initial scrutiny phase. The same researchers thoroughly studied and analyzed the full text of the research articles. While the third researcher validated and verified the results. The researchers assessed the quality of the research articles based on inclusion and exclusion criteria. We found satisfactory quality assessment results in the scrutiny process based on the aforesaid approach that includes, but is not limited to, strict adoption of inclusion and exclusion criteria, internal meetings to resolve minor variances between the researchers, and validation of the results. Moreover, we focused on the different vital factors like selection and assessment bias, related to threats to validity as well.

The execution of this phase decreased the papers to a total of 273 of 336 studies. Thus, our preliminary studies include 273 articles. We present our literature search strategy results in Figure 2. We added research articles of our preliminary studies in a reference manager tool to access, use, and manage references in our research work.

### 3.4 Analysis

In this section, we explain the fourth step of our methodology, i.e., the required data extraction from the research articles and presentation and organization of findings that formulate a framework to "fill-in" the identified research gaps.

We thoroughly analyzed the research articles from our preliminary studies. We extracted relevant information that includes definitions, types of big data, data actors and roles, and big data potential impact in PAs. Subsequently, we assembled and classified the extracted information and relevant research articles to answer the research questions RQ1-4. We used a spreadsheet program as a data extraction template to capture and record the information from the research articles.

We implemented the following steps to extract the results. First, we obtained the general information, like authors, title, publication year, type of publication. Second, studies were analyzed according to the above-mentioned inclusion and exclusion criteria. In case a study did not include in the SLR, then we did not extract the information from such research and excluded it from the analysis phase as well. In the third step, we arranged extracted data in the datasheet based on the critical aspects of our above-mentioned RQs.

To perform a detailed analysis, two researchers independently studied and analyzed the full text of the research articles. Both researchers compared their outcomes and found minor disagreements. Later, both researchers organized meetings to discuss and resolve their disagreements about text extraction. While the third researcher performed data extraction on a random sample, and then he verified and agreed with the outcomes. The analysis work reporting was based on the

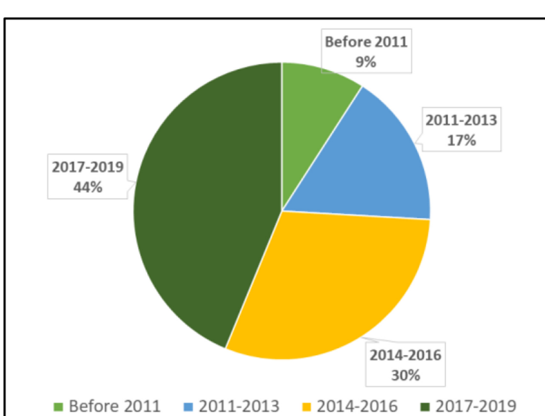


Fig. 3. Temporal distribution of research articles.

Table 1. Numbers of Research Articles per Research Question

| RQs Topics → | Definition | Types of big data | Actors & roles | Potential impact | Mixed mode |
|---|---|---|---|---|---|
| Total Papers → | 22 | 21 | 19 | 18 | 193 |

synthesized results. We used a descriptive synthesis approach to describe the results in a manner consistent with our RQs.

We portray an interesting descriptive statistic from our SLR process. Such statistics reflect the hype about the area of this study. The BDE is one of the new hot study areas among stakeholders, including governments. We observed that digital research libraries, particularly ACM and IEEE, promote BDE research works [37]. Our primary studies period distribution of research articles is given in Figure 3.

## 4 RESULTS

The last step of the above-mentioned SLR process explains the research results and offers a foundation for the study. Therefore, we carefully reviewed 273 research articles to extract relevant information. We arranged the information based on our four RQs. Table 1 indicates the number of research articles we found addressing each research question.

Under "Mixed mode" we grouped 193 research articles that discussed >1 of our RQs. We detailed answer to our RQs as below:

### 4.1 RQ1. Definition of GBDEs

The existing research studies present a heterogeneous theoretical foundation to define GBDEs. Such theories are often influenced by the socio-technical and value chain theories. These mixed theories are usually used in the literature to cover the theoretical gap as the big data field is in the early stages. Numerous business, research, and industry communities study the big data field [19, 20, 68]. Some definitions stay relevant to specific domains like humanitarian [32, 40] and personal data ecosystems [61]. These studies have a narrow perspective and focus on a certain notion with partial details [6, 31, 97]. We briefly describe here a few interesting literature definitions we found.

The first definition explains the humanitarian data ecosystem as a network of humanitarian actors, governments, private sector organizations, and affected communities in which they interact with each other to produce, collect and analyze digital data about vulnerable populations [32, 40]. Another definition describes the data ecosystem as a socio-technical complex network in which actors like organizations and individuals interact, collaborate to exchange and use data as the primary source to foster innovation and support new businesses [67]. In Reference [16], the data ecosystem is presented as a heterogeneous network of hardware, software, networking resources, human capital, industry applications, industry methodological techniques, social actors, and the new ideas and concepts those actors coin. Last, in Reference [83] the authors focus on the open data ecosystem and define it as complex of numerous actors, interacting with each other and performing open data related functions, i.e., to create, find, store, access, share, protect, preserve open data. We found more definitions of open data ecosystems in Reference [22, 35, 74, 99].

We identified three main concepts that are recurring in the BDE definitions. The three main concepts include "socio-technical network," "data functions," and "data value creation." We decided to organize our effort to propose a holistic definition based on these concepts. We present these concepts below.

The "*socio-technical network*" concept is about the interactions amongst socio-technical entities such as people, processes, technology, organizations, data, and infrastructure. This concept exists almost in all literature definitions. Some elements, like people and organizations, represent the social aspects, while others, like data models, data infrastructure, and data portals, represent the technological aspects. Though, different authors gave different labels to this concept in their definitions. Such labels include a "complex network of individuals and organizations," "interconnected human & technological resources," "complex interconnected, multilayered ecosystem," and "heterogeneous network of software, hardware, people, and processes." We assessed all these labels and assigned a unique label to this concept as a "socio-technical network." This also refers to certain characteristics of ecosystems, including but not limited to interdependencies among "people," "processes," "technology," "organizations," "data," "data infrastructure," "data analytics," and "data structure & model," "base registries," "data services," and "data portals" [35, 90].

"*Data functions*" refers to different elements, including data collection, data integration, analysis, data storage, sharing, access, use, data security and protection, publish, and data archive. These are essential steps to transform data/information into knowledge. This concept is also effective in identifying dataflows and work procedures for stakeholders [4], including what is shared between the so-called infomediaries [87].

*"Data value creation"* is about mining value from big data through its extensive use. We found a list of areas of data value creation from the literature including data-driven administration [95], public service delivery, data economy [22], innovation [22, 35, 68], new businesses [61, 99], data economy [22], transparency [22, 26, 49, 54, 90], public policies and strategies alignments [39, 88], open government [54], secure exchange of data [32], people's personal data management [29], government performance [90], democratic governance and political participation, trust in government, data reuse, and integration of public and private data [22].

During our review process, we observed that these three main concepts did not exist together in any single definition of the BDE. We present the occurrence of these three concepts in BDE definitions in Table 2.

For a vigorous definition of the GBDEs, we put together the above-mentioned three main concepts along with their characteristics, as found in the literature. The summary of the result is presented in Figure 4.

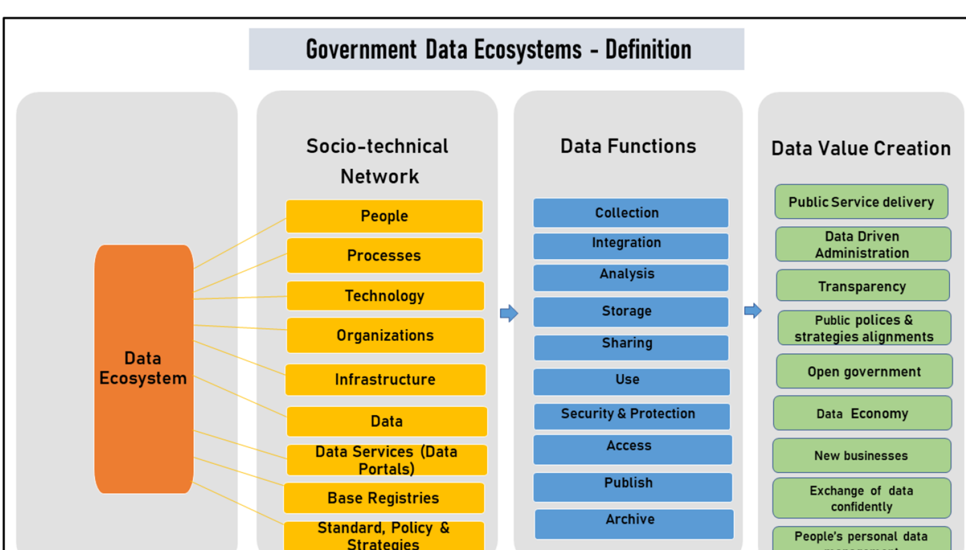


Fig. 4. Concepts of the GBDE.

Table 2. Occurrence of the Three Main Concepts in BDE Definitions

| S# | Literature Definitions References | Socio-technical Network | Data Functions | Data Value Creation |
|---|---|---|---|---|
| 1 | [68, 98, 67] | × | - | × |
| 2 | [25] | - | × | × |
| 3 | [99] | - | × | × |
| 4 | [35] | × | - | × |
| 5 | [23] | × | × | - |
| 6 | [22] | - | - | × |
| 7 | [74] | × | - | - |
| 8 | [72] | × | × | - |
| 9 | [14] | × | × | - |
| 10 | [87] | - | × | - |
| 11 | [97] | × | × | - |
| 12 | [16] | × | - | - |
| 13 | [31] | × | - | - |
| 14 | [42] | × | - | - |
| 15 | [32, 40] | × | × | - |
| 16 | [38] | × | - | - |
| 17 | [83] | × | × | - |
| 18 | [52] | × | - | × |
| 19 | [6] | - | × | × |

Hint: C1, Socio-technical network; C2, Data Functions; C3, Data Value Creation.

The first concept of our proposed definition provides sense to the readers that the GBDE consists of socio-technical elements, including organizations, require collaborative efforts to process the data big as per organization requirements [22]. The second concept explains a set of data functions through which organizations collect, integrate, filter, store, analyze, and visualize data for the future course of actions. The last concept emphasises that stakeholders should concentrate to better understand and assess the value of big data.

From the above, our proposed definition of GBDEs follows:

"A socio-technical network of people, processes, technology, infrastructure, data services, base registries standards & policies, processes, organizations, and resources jointly working to perform data functions such as data collection, integration, analysis, storage, sharing, use, security & protection and archiving to obtain value from big data through its extensive use to ensure better evidence-based policymaking, public services delivery, promote data-driven administration & open government, boost the data economy to benefit citizens, businesses, and government bodies."

### 4.2 RQ2. Types of Big Data

During the review process, we identified in the literature references to different types of big data. There are some studies [1, 15, 69, 94] focusing on typologies for big data, but each study considered different types of data based on the research perspective. In the literature, we did not find a holistic classification of types of big data. We did not get many research papers on the topic. Below, we present our own big data typology, for which we propose six dimensions. In Figure 5, we present a summary.

The description of each classification and its related types of (big) data is as below:

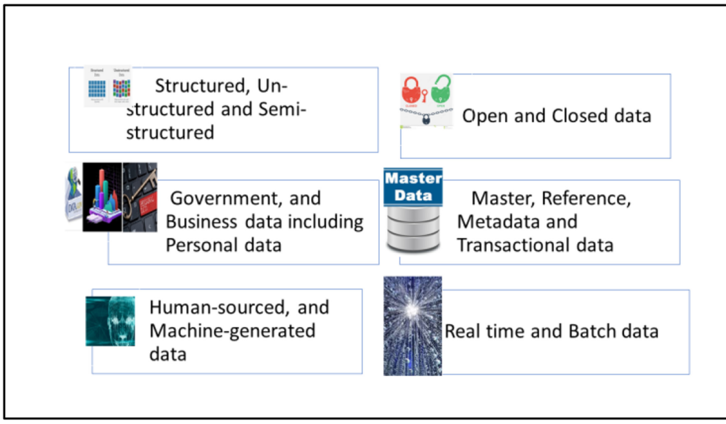


Fig. 5. Classification of Types of Big Data.

*4.2.1 Structured, Un-structured, and Semi-structured.* The first classification of big data types is effective in ascertaining how the incoming data need to be processed. It includes structured, unstructured, and semi-structured data types.

*Structured data* acquire highly structured schema, standard format, and arrangement [73]. Databases are extremely structured. Structure data are warehoused in the databases, retrieved, and processed in an arranged and thorough manner [73, 94]. The examples of structure data include library catalogs, phone directories, census records, and databases. Structured data account for some percentage of the total existing data. Structured data are self-describing utilized in computer programming for data collection, data cleansing, data visualization, and data analysis software tools [73, 94].

*Unstructured data:* Unstructured data, unlike structured data, are not abided by predefined layout, structure, data model, and do not match properly into relational tables [73, 94]. Consequently, it is difficult to transform or map unstructured data into the format needed for efficient and effective processing [73, 92]. Examples include images, audio, and video files. Scientific research data can also include unstructured data, also called “raw data,” e.g., gathered from scientists’ experiments [23]. Unstructured data has no explicit format in storage [94]. To further process structured data, data scientists need to pass it through several phases, which can be a time-consuming and tedious task [1].

*Semi-structured data* are not in the same layout as structured data, although it includes elements like semantic tags that make it easier to analyze. The semi-structured data include HTML [18, 94], XML, JSON documents, and content from NoSQL databases. Semi-structured data may be irregular or incomplete and have a structure that may change rapidly or unpredictably [73]. Unstructured and semi-structured data make up about 80% of the total existing data.

*4.2.2 Real-time Data and Batch Data.* It is critical to understand whether the data are analyzed in real time or batched for advanced study.

*Real-time big data* are created in real-time mode by systems [50]. Real-time examples of data include data from sensors, Twitter feeds, stock markets, traffic, bank ATMs, and radar weather data [50, 71]. A system that is centered on real-time data is identified as a real-time data management system. Such systems may offer advice and decision support to businesses, individuals, and governments on various affairs. For example, in Reference [50] a traffic navigation system is described based on real-time data. Navigation systems process real-time traffic data to propose optimal routes to their clients [23].

*Batch data* are gathered for some period before being available for further processing. In batch data systems, this time interval (lag) may vary from a few minutes to hours [71, 93]. Such data may include image processing, payroll, and billing data. The batch processing is an efficient way of handling big volumes of data. In batch data processing, data are loaded in huge batches at a specific interval [62]. Batch data processing is extensively used in the production of high-value, typically low-volume materials, including pharmaceuticals and microelectronics. The economic cost of lost process performance is usually high and has motivated extensive research in batch process monitoring, fault detection, and control. However, Wang et al. stated that data-driven methods have an essential role in this area [93]. High-availability distributed object oriented-Hadoop, an open-source software framework, is focused on batch data processing [2, 62].

*4.2.3 Human-sourced Data and Machine-generated Data.* This classification of big data type is beneficial to ascertain the scope from a business viewpoint.

*Human-generated data* are data produced from humans, e.g., records of human activities like work of art and books, photographs, audio, and video. Human-generated data can be analogue or totally digitized by default [69].

*Machine-generated data* are created by machines, e.g., computer/information systems, IoT, and mobile devices. Machine-generated data has size and speed far greater than human-generated data [30]. The IoT data sources include weather sensors, traffic sensors, and security/surveillance cameras, whereas computer systems data sources consist of computer logs and weblogs [7, 30] but also data from the execution of automated and computer-supported business processes [82]. A considerable rise in the number of sensors in the world results in growing data volumes generated from machines.

*4.2.4 Government, Business Data, Including Personal Data.* This classification comprises of big data types that are handled by public sector organizations. Though, such data correlates to citizens and businesses.

*Government data:* Digital data that relate to the government, and it can either be created or gathered by the PA. Government data are one of the critical assets of modern states [7]. The examples of government data include residents' social security numbers, criminal data, public scientific data, public health data, electoral rolls, and vehicle registration data. Government data often contains shared data that should be accessible from all or other specific public entities [7, 61]. Government entities use such shared data, amongst others, to attain greater efficiency in their organizations.

*Business data* are the information that is utilized to plan and run a business organization. The business data contain information about customers, places, products, and market trends. Patterns can be utilized to support the business, e.g., to forecast customer behavior [29]. Technical solutions support organizations to collect, warehouse, and track business data. For example, in the past, company salespeople used a Rolodex (business card holder) to store customer contacts, while nowadays, sales department staff use state-of-the-art CRM solutions to store customer data, requirements, behaviors, and customer gratification about brand products [45, 78, 98]. The quantity of business data is growing at an extreme speed. For example, Amazon is processing more than 35 transactions every second [78, 98].

Moiso and Minerva defined *Personal Data* as data regarding individuals, their conduct, and their actions [61]. Examples of personal data contain personal photos, credit card numbers, videos, chat, addresses. Government and business entities may have access and process personal data of both physical and legal persons [61]. Personal devices, like tablets, and smartphones, produce personal data. These data are assumed to become the "energy" or the "new currency" for the digital world.

Personal data about people and their activities also present new opportunities to people and organizations [61, 11]. The privacy of personal data is considered a great challenge nowadays. Therefore, several countries are adopting certain data protection laws, like the European General Data Protection Regulation, to manage the use of personal data. In parallel, risk-free data anonymization, in conjunction with good data governance practices, can also ensure better privacy protection for personal data [77].

*4.2.5 Open and Closed Data.* *Open data* are data that are freely accessible for use, reuse, and redistribution subject only, at most, to the necessity to attribute and share-alike [51, 60]. Governments produce and publish a large volume of open data; though, private firms could (or even should) also open a portion of their data [43]. Open data can propose new business openings for actors that offer data, for actors that utilize data, and for actors that construct innovative services and applications around the data [43]. Governmental open data portals comprise of different open datasets, for example, public procurement notices, economic and financial data, education, energy data, culture, and sports data, health, and public scientific research data.

*Closed data,* for our work, is used by government organizations and is not disclosed to third parties [73]. Closed data in businesses include, e.g., revenue data and product formula data, whereas closed data in the government sector include, e.g., employee service records, employee performance assessment reports, confidential and secret data [73]. Closed data access is limited to the data owner(s) and groups due to security limitations and relevant public policies. Public organizations need data security and data privacy measures for closed data [73]. Examples of such measures include the implementation of intrusion detection systems, firewalls, and enforcing access control policies [89].

*4.2.6 Master, Reference, Metadata, and Transactional Data.* In this part, we identify three important types of data for any organization.

*Master data* are a specific resource of primary business data utilized across numerous systems, applications, and processes [84]. Master data contain data that describes the highly relevant business entities on which the activities of an organization are based, like suppliers, counterparties, employees, or products [79]. Master data signify the critical transversal entities of the enterprises, particularly the business, that provides a context to the transactions [8, 79]. Examples of master data include *citizen master data* (a citizen's social security number, name, and address, etc.), *product master data* (product id#, product name, product unit), *customer master data* (customer id#, customer name, and address, etc.) [8, 34]. The master data facilitate the fundamental data entities of the government to attain a data-driven administration [79]. It is important to describe here that the master data discussion is often discussed under as "base registries" in PA.

*Reference data* are the set of allowable values to be utilized by other data fields [26]. Reference data are coded, semantically stable, comparatively static datasets shared by multiple constituencies like people, systems, and other master data domains [70]. Data users reuse reference data to obtains value [26, 51]. Errors in reference data affect the quality of master data and relevant transactional systems as well [70]. The examples of reference data include International Organization for Standardization country codes, Internet country code top-level domains, and International Telecommunication Union country phone codes [70].

*Metadata* are data about data. Metadata is a part of structured information that identifies, makes it easier to retrieve, track usage, and handle information resources [15]. For books, metadata examples include Book International Standard Book Number, Book title, Book ratings, and Book edition. Metadata should be produced along with the dataset in a regulated manner, published alongside the data, and revised whenever the dataset is updated [26, 51]. There are various types of metadata. Descriptive metadata includes elements that help to create and locate information resources. The structural metadata provides information about containers of data and indicates how compound objects are assembled. Administrative metadata provides information to facilitate the management of information resources [15, 44].

*Transactional data* are gained from business transactions, e.g., a company sells or a client purchases a product [3]. Transaction data can be exploited for business analytics and business intelligence. The transactional data contains some basics characteristics, like transaction ID and transaction date and time, to uniquely identify the transactions. Transactional data are more volatile than master data due to the frequent creation and changes in it. Usually, master data do not require to be altered or built with every transaction [3].

### 4.3 RQ3. Classification of Government Big Data Actors and Their Roles

Big data actor's typology signifies distinctive entities characterised by different types of relationships with data. We did not find an all-inclusive typology of big data actors in the literature. Authors identify different actors and assign divergent roles, e.g., References [38, 48, 61, 92, 98]. We propose a classification of government big data actors and their roles, summarized in Figure 6.

A detailed description of the above-mentioned classification follows below.

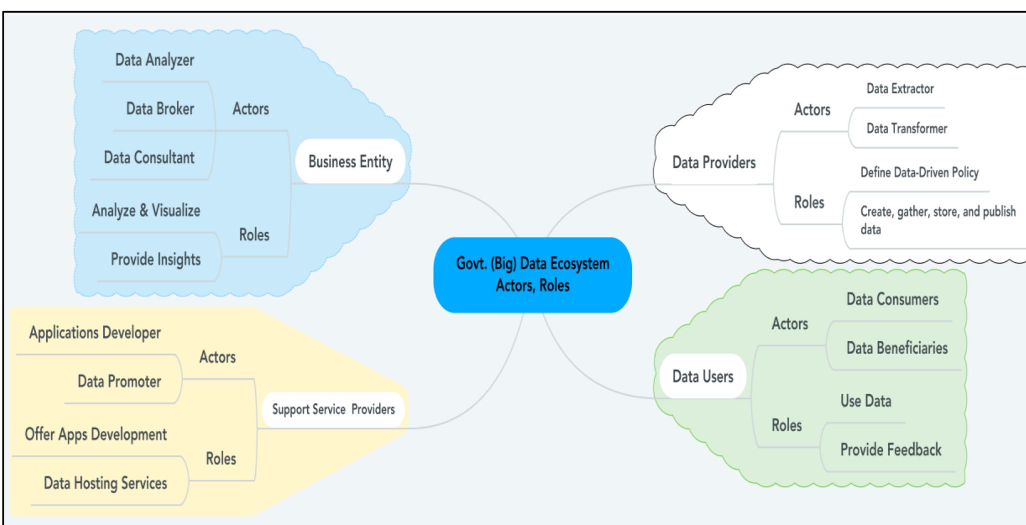


Fig. 6. Classification of government big data actors and their roles.

*4.3.1 C1: Data Publishers (Providers).* The data publishers offer data to other data actors of the GBDE [98]. The offered data include raw data, processed data, analyzed data, and so on.

The role of the data publishers is to create, collect data, store data, format data, and publish data. Data publishers with stakeholders can jointly take part in "feedback and discussion" activity. The motivation of the data publisher in PA is to promote better governance and to improve the quality of life [17, 32, 38, 61, 83]. Data providers and data users depend significantly on each other. The value of raw data could only attain when both actors work together [99]. Governmental data publishers also encourage the participation of citizens in governmental practices of decision making and policymaking [64, 98]. Data extractors and data transformers perform a role in collecting and shifting data in an appropriate format for a future course of action [54].

The data publishers may provide data free of charge or with some licenses that restrict the use of data for commercial or other purposes. The first category offers data for free, either without condition or with some license. Such a data license may restrict the use of data. For example, a Creative Commons license is a copyright permit that allows the free distribution of an otherwise copyrighted job [48, 98]. PAs [17, 32, 61], Non-Governmental Organizations (NGOs) [70], municipalities, state or federal government, and local government [80] usually belong to this category. They offer data to enhance the national economy, allowing businesses and citizens to utilize the data [85]. The second category generates revenue from selling access to data [85]. Such entities may offer a subset of their data as open data while guaranteeing accessibility only to paying users [78].

We present an example of a data publisher: the Helsinki Regional Transport Authority (HRTA), a data publisher, has developed a free Application Programming Interface (API) that provides access to information to data users. The data users, including private service providers, use this API to help customers with trip planning. Through HRTA API, data users can access different kinds of information, including public-transportation routes and timetables, service disruptions, and live data for vehicle location and tracking. Mobile Apps developers have used the HRTA API to create around 30 mobile trip-planning applications [54].

*4.3.2 C2: Data Users.* Data users consume data that are offered by data publishers in the GBDE. They may use information manually or with the support of data-based applications and services [43]. Governments, as active data users, are more capable of discovering the clients' needs and more likely to offer appropriate data and information [22].

Data users include data infomediaries or intermediaries. The infomediaries are the entities who process the raw data and add value to the data through data lifecycle phases like cleaning, analyzing, integrating, and visualizing the data. Additional data users may favor utilizing processed data or services and tools that are derived from raw data by infomediaries [98]. Data users include local community/individuals, public sector [61], NGOs, civil society [48], private sector, and academics [57].

Roles of data users include to discover, analyse, process, utilize data, give feedback, and perform R&D to examine new algorithms and data technologies [22, 83]. Data users' roles also include application developers who use data as part of the service [43], to search, combine, analyse, filter, visualize data [99]. Organizations and individuals may utilize this service [17]. The motives of the data users are to boost community welfare, business expansion, and to promote civic participation to enhance the quality of the data [22, 43].

*4.3.3 Data Business Entity.* The data business entities combine several publishers' data with their own expertise to offer data services to third parties. Such entities gain revenue by selling services to both data publishers and users [43, 56, 83,]. We observed different categories of data business entities including data application developers [43], data harmonizer [43, 48], data aggregator [38, 43], data enablers [56], data analyzer [54], data brokers [43, 83], data facilitators, data re-users [48], data marketplace companies, and data consultants [83]. Data aggregators mix and modify data and usually collect data from different sources [43, 61]. Data analyzers gather and analyze data [54]. Data harmonizers carry out standardization and homogenization of data. Data enablers offer solutions and services to data publishers to merge different types of data. Data brokers include data promoters, distributors, and matchmakers. Data brokers may sell personal data to other third parties without data owners' permission [66]. Data promoters acquire data and promote it to the other actors. Data distributors deliver communication and distribution data channels. Data matchmakers match data needs with the best available data sources [54]. Data facilitators assist with the interchange of data between the data publishers and data users [56]. Online data marketplace companies utilize smart applications. Such technological applications link buyers and sellers in an open, collaborative environment [83].

The following are some examples of data business entities: A data analyzer company "A" utilizes open government data and private data obtained from Finnish firms to generate credit ratings and other fiscal information for sale. There is another example of a data analyzer company "B" that analyzes business' financial data and represents an easy-to visualize, a tree-shaped image of their balance sheets [54]. The third example of data business entities includes data enablers companies' "Captricity" and "Xcential" that help PAs in transforming static documents into actionable data [56]. eBay, an e-commerce platform provider, Cream, and Uber, ride-hailing service providers, are examples of private online data marketplace companies [83].

*4.3.4 Data Support Service Providers.* The data support service provider offers technical support services to the other data actors in the GBDEs. Examples of such services include data storage, data hosting, development of mobile applications, and websites [43, 54]. There are various categories of data support service providers. These categories of service providers include cloud computing service providers, websites, and mobile applications development service providers, user-experience consultants, and so on [54, 43].

Our proposed classification of GBDE actors and their roles address the literature gap and provides a thorough set of actors and gives clarity in actors' roles, and their motivations.

### 4.4 RQ4. Impact of the GBDE

In this sub-section, we present an overview of the GBDE potential impact in different PA areas, including amongst others governance, policymaking, health, education, justice, budgeting, economy, agriculture, safety and security, transportation, and logistics.

*Financial Institutions:* Financial institutions are using big data and analytics to monitor unlawful financial market actions. Such public organizations are utilizing network analytics and natural language processing for monitoring purposes [63]. The financial institutions that include retail funds, big banks, and mutual funds are using big data analytics in high-frequency trading, decision support analytics before trading, predictive analytics, hazard investigation, risk management, and customer relationship [10, 63].

*Health:* Big data is booming in the healthcare area [33]. Health organisations utilize big data and analytics to detect early symptoms of various diseases, identification of medicine-usage irregularities, and enhanced tolerant security [12, 63]. Health care institutions analyze big data and extract useful insights to help experts to allocate resources effectively like nurses, clinicians, diagnostic machinery, and other resources [10]. PAs are utilizing big data and IoT technologies to enable telehealth and deliver continuous patient care, through smart devices, mobile Internet, and cloud services, in rural underserved areas of a county to enhance the effectiveness and quality of care [55]. Big data is also predicting the patient's hospitalization, discharge, and treatment time by analyzing the medical data and the patient's related records to improve the medical treatment effect [55].

*Education:* The public sector educational institutions are implementing big data to fine-tune the educational journey for the students to review, monitor, and track teacher's performance to improve education delivery to their communities. Big data in these institutions is also providing decision support to higher management to get excellence in the field of education [10]. PAs gather big data for education from various sources. Such big data sources include documents, emails, audit logs, Closed-Circuit Television footages, biometric devices, online surveys, student response systems, student and faculty feedback, and questionnaires [10, 63].

*Justice:* Big data is driving a trend toward behavioral optimization and "personalized law," in which legal decisions and rules are optimized for best outcomes and where the law is tailored to individual consumers based on analysis of past data [24]. PAs are utilizing big data and relevant technological tools to predict legal costs and legal case outcomes, manage data for regulatory compliance, and reduce document review costs [24, 63]. In public sector legal organizations, legal experts are also using sophisticated predictive big data analytics software to analyze data and compare it to the facts of legal cases [24]. Moreover, PAs are using big data algorithms to mine prior precedents or other relevant data for correlations between variables, such as by finding common factors in a judge's previous decisions that are predictive of future outcomes.

*Budgeting and Policymaking:* The public sector financial organizations are utilizing big data to improve the reliability of budgeting estimates and enable complex real-time scenarios [39]. In recent years, the involvement of parliamentarians with budgeting has grown. Usually, parliamentarians have limited expertise to formulate a well-informed budget as compared to the senior government executives. However, this capacity is essential for well-informed budgeting and actual execution. The senior executives support the parliamentarians in the budgeting and policymaking tasks. The big data and data analytics help the parliamentarians, senior executives, and other stakeholders to formulate well-informed budgeting, examine the policy documents and shape actual national policies as per the needs of communities [39, 78].

GBDE helps policymakers to reduce the time frame and boost the evidence-based policy decisions by using advanced predictive analytics methodologies and scenario techniques. In governments, politicians use big data platforms and analytical tools to estimate public opinion through analyzing social media and to produce census data daily. Governments, particularly in advanced countries across the globe, are implementing big data solutions in parliaments to create a data-driven culture within their premises. Such smart parliaments have numerous big data users like Senators, Members, parliamentary committees, and legislative departments. Parliament big data contains the record of parliamentary debates, votes information, notice papers, media releases, committee hearings, and reports that they publish in print and electronic media. Smart parliament BDE also contains relevant press and academic articles, media programs that inform and commentary on the workings of the politicians, parliament, political parties, and policies [53]. The parliamentarians can use such parliamentary big data to hold a healthy deliberation on well-informed budgeting, national policies and subsequently decide to approve a budget and evidence-based policies with consensus to find out possible solutions to the critical national issues [39].

*Economic Development and Agriculture:* Public sector organizations and business organizations, as part of the smart and wider digital and knowledge economies, are leading innovation in the big data space and have a vital role to play in supporting economic growth. The public sector organizations are creating partnerships with big data businesses for economic development [6, 72]. Globally, PAs can utilize big data in various public sectors, particularly in the agriculture sector, to boost productivity, food security, and farmer incomes at the same time. The government agriculture organizations staff and farmers are using big data technologies to access real-time data to obtain the functioning status of farm machinery, historical weather patterns, topography, and crop performance [71].

*Safety and Security:* The government safety, and security-related departments, like police departments, are using big data and analytics for individual criminal-behaviors prevention, detection of organized crime, and corrections optimization. The above-mentioned departments are using big data and data analytics to perform analysis of crime history data to forecast and predict incidents and potential impacts for a more proactive response. Police departments can maximize the use of IoTs to collect and share big data amongst relevant stakeholders like police, citizens [11]. The mixing of heterogeneous data from multiple sources of data also provides reliable and up-to-date information and advanced safety and security services to the citizens

*Transportation and Logistics:* PAs generate transportation big data and use it to control traffic, optimize infrastructure, to plan the route, to develop intelligent transport systems, and to enhance road safety [63, 55]. Citizens use big data to plan routes to save on fuel and time, and in tourism, the industry can use it for tour arrangements [36].

The potential impacts of [big] data in the *other areas of PA* include bureaucracy, census, smart voting, open big data government and military operations [10, 24, 39, 91].

## 5 CONCLUSION AND FUTURE WORK

In the era of the data-driven world, the GBDE receives interest and attention from researchers and professionals as it is the essential handler and enabler for the data-driven government. In the GBDE, data actors interact with each other and perform various data functions. Still, GBDE fundamental elements are not well defined. As a result, its meaning is yet tenuous, and this impedes progress and evolution. We conducted a SLR regarding the fundamental elements of the GBDE. We have thoroughly organised the fundamental elements of GBDE under a framework to establish a common ground to discuss and research the subject area. Such elements include (a) a definition for the GBDE, (b) a classification of government big data actors and their roles, (c) a classification of government big data types of big data, and (d) the impact of big data in core PAs sectors. To the best of our knowledge, it is the first academic study that categorically addresses the above elements of the GBDE.

### 5.1 Research Implications

In view of the above-mentioned findings, we proposed the following research implications for the scholars and practitioners:

**Benefits to the Research Community:**

- Our proposed framework, which addresses the above-mentioned four RQs, may aid the research community in their current and future research initiatives related to the GBDE.
- This study offers a common framework to bring the research community under a unique and consistent framework umbrella to cultivate common ground, starting from the definition of GBDEs, identifying types of big data, classifying data actors, and impacted areas of PAs related to respective governmental ecosystems.

**Benefit to Entrepreneurs:**

- This study may offer insights regarding the essential elements of GBDE to entrepreneurs so that they can evaluate new business opportunities based on innovative ideas in big data solutions and services.

**Benefits to the Practitioners:**

- The elements of the proposed framework could be helpful for the practitioners to understand GBDE (definition element), comprehend data topology, as one of the prime factors, to build or select best-fit big data solutions, get a clue about their own and other data actors roles and spot critical areas of PA wherein big data has potential impacts for their future course of actions.

- The elements of the GBDE, particularly data actors and their roles, as an information tool, may be useful for the public sector organizations to develop or modify their strategic measures to become a data-driven organization.

**Benefits for the Public Administrations**

- Our contribution regarding the impact of big data in PA may be helpful to create awareness and understanding about the critical areas of PA wherein big data may have potential effects.
- PAs may consider the proposed areas and related big data initiatives in their environment to manage public sector big data and information as an asset to boost organizational, operational efficiencies, data-driven decision making, civic participation in the formation of national policies, and to improve services.

### 5.2 Limitations

We identify the following main limitations of our work:

- We found research articles on BDEs in areas such as open government, scientific research, business, and the semantic web. But we did not find many research articles explicitly on the GBDE. Therefore, we borrowed concepts for the GBDE from the existing literature in "neighboring" areas.
- We adopted SLR as a research method. In this method, the selection of digital research libraries is a challenging task. We selected and investigated four digital research libraries; however, other research libraries may contain relevant research articles about the GBDE.
- The formulation of the search string to search literature has significant impacts on the outcomes of systematic research studies. We have done our best to handle associated hazards in the formulation of the search strings. However, the study is yet restricted by the selected search string.

### 5.3 Future Work

- This study mainly contributes to the theoretical realm and requires a practical verification to bridge the gap between academic rigor and industry relevance. To manage the practical area, we propose that the outcomes of this study need to be tested on its application in the government sector organization utilizing a case study.
- In this study, we researched the GBDE as a new field of growing significance. However, some critical areas need special attention from the relevant research community. An example of such areas includes theories about the BDE. Such theories offer a conceptual basis for future research about how to develop and evolve a BDE and to build a common understanding of it. Therefore, the research community may consider BDE theories as a priority topic for future research work.
- Last, we intend to extend our proposed framework for the GBDEs to include additional aspects like the big data lifecycle, the components of GBDEs, i.e., data infrastructure, data analytics, data services (data portals), standards, policy and strategy, base registries, data structures, and models.

## REFERENCES

[1] Mohammad Fikry Abdullah and Kamsuriah Ahmad. 2013. The mapping process of un-structured data to structured data. In *Proceedings of the International Conference on Research and Innovation in Information Systems (ICRIIS'13)*. 151–155. DOI: https://doi.org/10.1109/ICRIIS.2013.6716700

[2] Umar Aftab and Ghazanfar Farooq Siddiqui. 2019. (Big) Data augmentation with data warehouse: A survey. In *Proceedings of the IEEE International Conference on Big Data (Big Data'18).* 2785–2794. DOI: https://doi.org/10.1109/BigData.2018.8622206

[3] Philip Woodal Alexander Borek, Ajith K. Parlikad, and Jela Webb. 2014. *Total Information Risk Management—Maximizing the Value of Data and Information Assets.* Front Matter. DOI: https://doi.org/https://doi.org/10.1016/C2012-0-00446-2

[4] Suzie Allard. 2012. DataONE: Facilitating escience through collaboration. *J. eSci. Libr.* 1, 1 (2012), 4–17. DOI: https://doi.org/10.7191/jeslib.2012.1004

[5] Arlene Fink. 2010. *Conducting Research Literature Reviews - from the Internet to paper* (3rd ed.). SAGE, Thousand Oaks, California.

[6] Judie Attard, Fabrizio Orlandi, and Soren Auer. 2017. (Big) data value networks: Enabling a new data ecosystem. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'16).IE* 453–456. DOI: https://doi.org/10.1109/WI.2016.0073

[7] Judie Attard, Fabrizio Orlandi, and Sören Auer. 2016. Value creation on open government data. In *Proceedings of the Annual Hawaii International Conference on System Sciences.* 2605–2614. DOI: https://doi.org/10.1109/HICSS.2016.326

[8] Ehsan Baghi, Simon Schlosser, Verena Ebner, Boris Otto, and Hubert Oesterle. 2014. Toward a decision model for master data application architecture. In *Proceedings of the Annual Hawaii International Conference on System Sciences.* 3827–3836. DOI: https://doi.org/10.1109/HICSS.2014.475

[9] Michael J. Becker. 2014. The consumer data revolution: The reshaping of industry competition and a new perspective on privacy. *J. Direct Data Digit. Mark. Pract.* 15, 3 (2014), 213–218. DOI: https://doi.org/10.1057/dddmp.2014.3

[10] Anjum Zameer Bhat and Imran Ahmed. 2016. Big data for institutional planning, decision support and academic excellence. In *Proceedings of the 3rd MEC International Conference on Big Data and Smart Computing (ICBDSC'16).* 118–122. DOI: https://doi.org/10.1109/ICBDSC.2016.7460353

[11] Evmorfia Biliri, Panagiotis Kokkinakos, Ariadni Michailitsi, Dimitris Papaspyros, John Tsapelas, Spiros Mouzakitis, Sotiris Koussouris, Fenareti Lampathaki, Yury Glickman, and Fabian Kirstein. 2018. (Big) data analytics in public safety and personal security- Challenges and potential. In *Proceedings of the International Conference on Engineering, Technology, and Innovation and International Technology Managemant Conference (ICE/ITMC'17)* 1382–1386. DOI: https://doi.org/10.1109/ICE.2017.8280043

[12] Razan F. Binyousef, Aboelkhair M. Al-Gahmi, Zahid Rahman Khan, and Elham Rawah. 2017. A rare case of erdheim-chester disease in the breast. *Ann. Saudi Med* 37, 1 (2017), 79–83. DOI: https://doi.org/10.5144/0256-4947.2017.79

[13] Michael L. Brodie. 2010. Data Integration at scale: From relational data integration to information ecosystems. In *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications.* 1–3. DOI: https://doi.org/10.1109/aina.2010.184

[14] Sharon Buteau, Preethi Rao, Anshuman Kr M., and Vigneshraja K. 2018, Developing a framework to assess socio-economic value of open (Big) data in India. In *Proceedings of the 14th International Symposium on Open Collaboration (OpenSym'18).* 1–6. DOI: https://doi.org/10.1145/3233391.3233532

[15] Liam J. Caffery, David Clunie, Clara Curiel-Lewandrowski, Josep M. H. Peter Soyer, and Allan C. Halpern. 2018. Transforming dermatologic imaging for the digital era: Metadata and standards. *J. Digit. Imag.* 31, 4 (2018), 568–577. DOI: https://doi.org/10.1007/s10278-017-0045-8

[16] Bongsug Chae. 2015. (Big) Data and IT-enabled services: Ecosystem and coevolution. *IT Prof.* 17, 2 (2015), 20–25. DOI: https://doi.org/10.1109/MITP.2015.17

[17] Sumandro Chattapadhyay. 2014. Access and use of government data by research and advocacy organisations in India. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV'14).* ACM, 361–364. DOI: https://doi.org/10.1145/2691195.2691262

[18] Yi Chen, Wei Wang, Ziyang Liu, and Xuemin Lin. 2009. Keyword search on structured and semi-structured data. In *Proceedings of the International Conference of the Special Interest Group on the Management of and the Symposium on Principles of Database Systems (SIGMOD-PODS'09).* 1005–1009. DOI: https://doi.org/10.1145/1559845.1559966

[19] Ching Chyi Lee and Jie Yang. 2000. Knowledge value chain. *J. Manag. Dev* 19, 9 (2000), 783–794. DOI: https://doi.org/10.1108/02621710010378228

[20] Chris W. Clegg. 2000. Socio-technical principles for system design. *Appl. Ergon.* 31, 5 (2000), 463–477. DOI: https://doi.org/10.1016/S0003-6870(00)00009-0

[21] L. V. Cooper and H. Hedges. 2009. Research synthesis as a scientific process. In *The Handbook of Research Synthesis and Meta-Analysis.*

[22] Sharon S. Dawes, Lyudmila Vidiasova, and Olga Parkhimovich. 2016. Planning and designing open government (big) data programs: An ecosystem approach. *Gov. Inf. Q* 33, 1 (2016), 15–27. DOI: https://doi.org/10.1016/j.giq.2016.01.003

[23] Yuri Demchenko, Cees De Laat, and Peter Membrey. 2014. Defining architecture components of the (Big) data ecosystems. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS'14).* 104–112.

[24] Caryn Devins, Teppo Felin, Stuart Kauffman, and Roger Koppl. 2017. The law and big data. *Cornell J. Law Publ. Pol.* 27, 2 (2017), 357–413.

[25] Li Ding, Timothy Lebo, John S. Erickson, Dominic Difranzo, Gregory Todd Williams, Xian Li, James Michaelis, Alvaro Graves, Jin Guang Zheng, Zhenning Shangguan, Johanna Flores, Deborah L. McGuinness, and James A. Hendler. 2011. TWC LOGD—A portal for linked open government data ecosystems. *J. Web Semant.* 9, 3 (2011), 325–333. DOI: https://doi.org/10.1016/j.websem.2011.06.002

[26] Li Ding, Vassilios Peristeras, and Michael Hausenblas. 2012. Linked open government data [guest editors' introduction]. *IEEE Intell. Syst.* 27, 3 (2012), 11–15. DOI: https://doi.org/10.1109/mis.2012.56

[27] Aleksander Fabijan, Pavel Dmitriev, Helena Holmstrom Olsson, and Jan Bosch. 2017. The evolution of continuous experimentation in software product development: From data to a data-driven organization at scale. In *Proceedings of the 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE'17)*. 770–780. DOI: https://doi.org/10.1109/ICSE.2017.76

[28] Ronan Fahy, Joris Van Hoboken, and Nico Van Eijk. 2018. Data privacy, transparency and the data-driven transformation of games to services. In *Proceedings of the 2018 IEEE Games, Entertainment and Media Conference (GEM'18)*. 136–146. DOI: https://doi.org/10.1109/GEM.2018.8516441

[29] Shaokun Fan, Raymond Y. K. Lau, and J. Leon Zhao. 2015. Demystifying (Big) data analytics for business intelligence (BI) through the lens of marketing mix. *Big Data Res.* 2, 1 (2015), 28–32. DOI: https://doi.org/10.1016/j.bdr.2015.02.006

[30] Nicolas Ferry, German Terrazas, Per Kalweit, Arnor Solberg, Svetan R., and Dirk Weinelt. 2017. towards a (big) data platform for managing the machine generated (big) data in the cloud. In *Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN'17)*. 263–270. DOI: https://doi.org/10.1109/INDIN.2017.8104782

[31] Robert L. Grossman. 2018. data lakes, commons and clouds: A review for analyzing and sharing genomic data. 1–5. arXiv:1809.01699. Retrieved from https://arxiv.org/abs/1809.01699.

[32] Elise Haak, Jolien Ubacht, Marc Van Den Homberg, Scott Cunningham, and Bartel Van Den Walle. 2018. A framework for strengthening data ecosystems to serve humanitarian purposes. In *ACM International Conference Proceedings Series* (2018). DOI: https://doi.org/10.1145/3209281.3209326

[33] M. Hadzic and T. S. Dillon. 2008. Application of digital ecosystems in health domain. *2008 2nd IEEE Int. Conf. Digit. Ecosyst. Technol. IEEE-DEST 2008* (2008), 543–547. DOI: https://doi.org/10.1109/DEST.2008.4635222

[34] Faizura Haneem, Rosmah Ali, Nazri Kama, and Sufyan Basri. 2017. Resolving data duplication, inaccuracy and inconsistency issues using master data management. In *Proceedings of the International. Conference on Research and. Innovation in Information Systems (ICRIIS'17)*. 1–6. DOI: https://doi.org/10.1109/ICRIIS.2017.8002453

[35] Teresa M. Harrison, Theresa A. Pardo, and Meghan Cook. 2012. Creating open government ecosystems: A research and development agenda. *Futur. Internet* 4, 4 (2012), 900–928. DOI: https://doi.org/10.3390/fi4040900

[36] Ibrahim Abaker Targio Hashem, Victor Chang, Nor Badrul Anuar, Kayode Adewole, Ibrar Yaqoob, Abdullah Gani, Ejaz Ahmed, and Haruna Chiroma. 2016. The roles of (big) data in smart city. *Int. J. Inf. Manage* 36, 5 (2016), 748–758. DOI: https://doi.org/10.1016/j.ijinfomgt.2016.05.002

[37] Bingsheng He, Yunji Chen, and Jingren Zhou. 2019. Guest editors introduction: Special issue on big data systems on emerging architectures. *IEEE Trans. Big Data* 5, 1 (2019), 1–3. DOI: https://doi.org/10.1109/tbdata.2019.2898067

[38] Maximilian Heimstädt. 2014. The institutionalization of digital openness. In *Proceedings of the International Symposium on Open Collaboration (OpenSym'14)* ACM Press, New York, NY, 1–2. DOI: https://doi.org/10.1145/2641580.2641626

[39] Johann Höchtl, Peter Parycek, and Ralph Schöllhammer. 2016. Big data in the policy cycle: Policy decision making in the digital era. *J. Organ. Comput. Electron. Commer.* 26, 1–2 (2016), 147–169. DOI: https://doi.org/10.1080/10919392.2015.1125187

[40] Marc Van Den Homberg, Jannis Visser, and Maarten Van Der Veen. 2017. Unpacking data preparedness from a humanitarian decision making perspective: Toward an assessment framework at subnational level. In *Proceedings of the International ISCRAM Conference.* 2–13.

[41] Anne Immonen and Jarmo Kalaoja. 2019. Requirements of an energy data ecosystem. *IEEE Access* 7, (2019), 111692–111708. DOI: https://doi.org/10.1109/access.2019.2933919

[42] Anne Immonen, Eila Ovaska, and Tuomas Paaso. 2018. Towards certified open (big) data in digital service ecosystems. *Softw. Qual. J.* 26, 4 (2018), 1257–1297. DOI: https://doi.org/10.1007/s11219-017-9378-2

[43] Anne Immonen, Marko Palviainen, and Eila Ovaska. 2014. Requirements of an open data based business ecosystem. *IEEE Access* 2, (2014), 88–103. DOI: https://doi.org/10.1109/ACCESS.2014.2302872

[44] Jenn Riley. 2017. Understanding Metadata: What Is Metadata. Retrieved from http://www.niso.org/apps/group_public/download.php/17446/Understanding Metadata.pdf.

[45] Asif Uddin Khan and Bikram Kesari Ratha. 2016. Business data extraction from social networking. In *Proceedings of the 2016 3rd International Conference on Recent Advances in Information Technology (RAIT'16*). 651–656. DOI: https://doi.org/10.1109/RAIT.2016.7507976

[46] Kitchenham. 2007. Guideline for Performing Systematic Literature Reviews (SLRs) in Software Engineering. Retrieved from https://s3.amazonaws.com/academia.edu.documents/35830450/2_143465389588742151.pdf?response-content-disposition=inline%3B filename%3DSystematic_Literature_Reviews_SLR.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20191206%2Fu.

[47] Barbara Kitchenham, Rialette Pretorius, David Budgen, O. Pearl Brereton, Mark Turner, Mahmood Niazi, and Stephen Linkman. 2010. Systematic literature reviews in software engineering-a tertiary study. *Inf. Softw. Technol.* 52, 8 (2010), 792–805. DOI: https://doi.org/10.1016/j.infsof.2010.03.006

[48] Victoria Köster and Gustavo Suárez. 2016. Open data for development: Experience of Uruguay. In *ACM International Conference Proceeding Series.* 207–210. DOI: https://doi.org/10.1145/2910019.2910060

[49] Dmitrij Koznov, Olga Andreeva, Uolevi Nikula, Andrey M., Dmitry Muromtsev, and Irina Radchenko. 2016. A survey of open government data in russian federation open data movements in different. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (Ic3k'16).* 173–180. DOI: https://doi.org/10.5220/0006049201730180

[50] Kam-yiu Lam, Alan Kwan, and Krithi Ramamritham. 2002. RTMonitor : Real-time data monitoring using mobile agent technologies. In *Proceedings of the 28th International Conference on Very Large Data Bases* (2002).

[51] Deirdre Lee. 2014. Building an open data ecosystem: an Irish experience. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV'14).* ACM, New York, NY, USA, 351–360. DOI: https://doi.org/10.1145/2691195.2691258

[52] Marco Iansiti and Roy Levien. 2004. *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability.* Harvard Business School Press.

[53] Senior Law Librarian, Bills Digest Section, and Australian Parliamentary Library. 2017. A sea of 'parliamentary big data': Collaborative approaches to its navigation by the australian parliamentary library research service. 1–8.

[54] Juho Lindman, Tomi Kinnari, and Matti Rossi. 2016. Business roles in the emerging open-data ecosystem. *IEEE Softw.* 33, 5 (2016), 54–59. DOI: https://doi.org/10.1109/MS.2015.25

[55] Xingye Lv. 2019. The big data impact and application study on the like ecosystem construction of open internet of things. *Cluster Comput.* 22, S2 (March 2019), 3563–3572. DOI: https://doi.org/10.1007/s10586-018-2206-z

[56] Gustavo Magalhaes, Catarina Roseira, and Laura Manley. 2014. Business models for open government data. In *ACM International Conference Proceedings Series.* 365–370. DOI: https://doi.org/10.1145/2691195.2691273

[57] José María and Cavanillas Edward. 2016. *New Horizons for a Data-Driven Economy.* Library of Congress. DOI: https://doi.org/10.1007/978-3-319-21569-3

[58] Somnath Mazumdar, Daniel Seybold, Kyriakos Kritikos, and Yiannis Verginadis. 2019. *A Survey on Data Storage and Placement Methodologies for Cloud-Big Data Ecosystem.* Springer International Publishing. DOI: https://doi.org/10.1186/s40537-019-0178-3

[59] Amel Ben Hadj Salem Mhamdia. 2013. Performance measurement practices in software ecosystem. *Int. J. Product. Perform. Manag.* 62, 5 (2013), 514–533. DOI: https://doi.org/10.1108/IJPPM-09-2012-0097

[60] Durgaprasad Misra, Alka Mishra, Sunil Babbar, and Varun Gupta. 2017. Open Government Data Policy and Indian Ecosystems. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance (ICEGOV'17).* ACM, New York, NY, USA, 218–227. DOI: https://doi.org/10.1145/3047273.3047363

[61] Corrado Moiso and Roberto Minerva. 2012. Towards a user-centric personal data ecosystem The role of the bank of individuals' data. In *Proceedings of the 16th International Conference on Intelligence in Next Generation Networks.* IEEE, Berlin, Germany, 202–209. DOI: https://doi.org/10.1109/ICIN.2012.6376027

[62] P. Muthulakshmi and S. Udhayapriya. 2018. A survey about (big) data issues and challenges. *Int. J. Comput. Sci. Eng.* 6, 6 (2018), 1238–1244. DOI: https://doi.org/10.26438/ijcse/v6i6.12381244

[63] Kirtida Naik and Abhijit Joshi. 2017. Role of big data in various sectors. In *Proceedings of the International Conference on IoT Socal, Mobile, Analytics and Cloud (I-SMAC 2017).* 117–122. DOI: https://doi.org/10.1109/I-SMAC.2017.8058321

[64] NIST Big Data Public Working Group: Definitions and Taxonomies Subgroup. 2015. NIST Big Data Interoperability Framework: Volume 2, Big Data Taxonomies.*2*, 31. DOI: https://doi.org/10.6028/NIST.SP.1500-2

[65] Jabu Mtsweni Nobubele, Angel Shozi. (Big) data privacy and security—A systematic analysis of current and future challenges. 1–9.

[66] Hyeontaek Oh, Sangdon Park, Gyu Myoung Lee, Hwanjo Heo, and Jun Kyun Choi. 2019. Personal data trading scheme for data brokers in iot data marketplaces. *IEEE Access* 7, (2019), 40120–40132. DOI: https://doi.org/10.1109/ACCESS.2019.2904248

[67] Marcelo Iury S. Oliveira, Lairson Emanuel R. A. Oliveira, Marlos G. Ribeiro Batista, and Bernadette Farias Lóscio. 2018. Towards a meta-model for data ecosystems. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age (dg.o'18).* 1–10. DOI: https://doi.org/10.1145/3209281.3209333

[68] Marcelo Iury S. Oliveira, Glória de Fátima Barros Lima, and Bernadette Farias Lóscio. 2019. Investigations into Data Ecosystems: A systematic mapping study. *Knowl. Inf. Syst.* 61, 2 (November 2019), 589–630. DOI: https://doi.org/10.1007/s10115-018-1323-6

[69] Alice J. Otoole and P. Jonathon Phillips. 2017. Five principles for crowd-source experiments in face recognition. In *Proceedings of the 12th IEEE International Conference Automatic Face Gesture Recognition (FG'17)*. 735–741. DOI: https://doi.org/10.1109/FG.2017.146

[70] Boris Otto, Kai M. Hüner, and Hubert Österle. 2012. Toward a functional reference model for master data quality management. *Inf. Syst. E-bus. Manag.* 10, 3 (2012), 395–425. DOI: https://doi.org/10.1007/s10257-011-0178-0

[71] Gautam Pal, Gangmin Li, and Katie Atkinson. 2018. Big data real time ingestion and machine learning. In *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing (DSMP'18)* 25–31. DOI: https://doi.org/10.1109/DSMP.2018.8478598

[72] Mark A. Parsons, Øystein Godøy, Ellsworth Ledrew, Taco F. De Bruin, Bruno Danis, Scott Tomlinson, and David Carlson. 2011. A conceptual framework for managing very diverse (big) data for complex, interdisciplinary science. *J. Inf. Sci* 37, 6 (2011), 555–569. DOI: https://doi.org/10.1177/0165551511412705

[73] Ajeet Ram Pathak, Manjusha Pandey, and Siddharth Rautaray. 2018. Construing the big data based on taxonomy, analytics and approaches. *Iran J. Comput. Sci* 1, 4 (2018), 237–259. DOI: https://doi.org/10.1007/s42044-018-0024-3

[74] Antti Poikola, Petri Kola, and K. A. Hintikka. 2011. *Public data-an introduction to opening information resources.* Edita Prima Oy, Helsinki, Finland. Retrieved from https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78201/Public_data_-_an_introduction_to_opening_information_resources.pdf?sequence=1.

[75] Shital Poojary and Jimish Deepak Bagadia. 2014. Reviewing literature for research: Doing it the right way. *Ind. J. Sex. Transm. Dis* 35, 2 (2014), 85–91. DOI: https://doi.org/10.4103/0253-7184.142387

[76] Marco Pospiech and Carsten Felden. 2013. A descriptive big data model using grounded theory. In *Proceedings of the 16th IEEE International Conference on Computer Science Engineering (CSE'13)*. 878–885. DOI: https://doi.org/10.1109/CSE.2013.132

[77] Artur Potiguara Carvalho, Fernanda Potiguara Carvalho, Edna Dias Canedo, and Pedro Henrique Potiguara Carvalho. 2020. Big data, anonymisation and governance to personal data protection. In *ACM International Conference Proceedings Series.* 185–195. DOI: https://doi.org/10.1145/3396956.3398253

[78] Erico Przeybilovicz and Maria Alexandra Cunha. 2018. Open government data programs. In *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age (dgo'18)*, ACM Press, New York, New York, USA, 1–2. DOI: https://doi.org/10.1145/3209281.3209346

[79] Nanik Qodarsih, Satrio Baskoro Yudhoatmojo, and Achmad Nizar Hidayanto. 2019. Master data management maturity assessment: A case study in the supreme court of the republic of indonesia. In *Proceedings of the 2018 6th International Conference Cyber IT Service Management (CITSM'18)*. DOI: https://doi.org/10.1109/CITSM.2018.8674373

[80] Rahul Kapoor Ron Adner. 2010. Value creation in innovation ecosystems—How the structure of technological interdependence affects firm performance in new technology generations. *Strateg. Manag. J.* 31, 1 (2010), 306–333. DOI: https://doi.org/10.1002/smj.821

[81] Tatiana Lukoianova and Victoria L. Rubin. 2014. Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Adv. Classif. Res.* Online 24, 1 (January 2014), 4–15. DOI: https://doi.org/10.7152/acro.v24i1.14671

[82] Khurram Shahzad and Jelena Zdravkovic. 2011. Towards goal-driven access to process warehouse: Integrating goals with process warehouse for business process analysis. In *Proceedings of the International Conference on Research Challenges in Information Science* (2011). DOI: https://doi.org/10.1109/RCIS.2011.6006861

[83] Goran Smith, Hosea Ayaba Ofe, and Johan Sandberg. 2016. Digital service innovation from open data: Exploring the value proposition of an open data marketplace. In *Proceedings of the Annual Hawaii International Conference on System Science.* 1277–1286. DOI: https://doi.org/10.1109/HICSS.2016.162

[84] Heather A. Smith and James D. McKeen. 2008. Developments in Practice XXX: Master Data Management: Salvation Or Snake Oil? *Commun. Assoc. Inf. Syst.* 23, 1 (2008), 63–72. DOI: https://doi.org/10.17705/1CAIS.02304

[85] Ken Smith, Len Seligman, Arnon Rosenthal, Chris Kurcz, Mary Greer, Catherine MacHeret, Michael Sexton, and Adric Eckstein. 2014. "Big metadata": The need for principled metadata management in big data ecosystems. In *Proceedings of the 3rd Workshop on Data Analytics in the Cloud (DanaC'14)*. 21–24. DOI: https://doi.org/10.1145/2627770.2627776

[86] Statista Organization. Internet of Things (IoT) Connected Devices Installed Base Worldwide from 2015 to 2025 (in billions). Retrieved from https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[87] Rufus Pollock. 2011. Building the (Open) Data Ecosystem – OKF blog. 1–8. Retrieved September 15, 2020 from https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/.

[88] Patrick O'Donovan T. D. 2018. *PUBLIC SERVICE DATA STRATEGY - Online Document.* Retrieved September 15, 2020 from https://data.gov.ie/uploads/page_images/2019-01-03-110200.740673Public-Service-Data-Strategy-2019-2023.pdf.

[89] Anucha Tungkasthan, Pitaya Poompuang, and Sarayut Intarasema. 2019. Government data sharing framework based on DIKW hierarchy model. In *Proceedings of the International Conference on ICT Knowledge Engineering.* 8–11. DOI: https://doi.org/10.1109/ICTKE47035.2019.8966872

[90] Barbara Ubaldi. 2013. OECD iLibrary | Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. 1–61. Retrieved September 15, 2020 from https://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en.

[91] Karishma Varshney, Rahul Johari, and R. L. Ujjwal. 2018. BiDuR: Big data analysis in UAF using R-voting system. In *Proceedings of the 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO'17)*. 333–337. DOI: https://doi.org/10.1109/ICRITO.2017.8342447

[92] Syed Iftikhar Hussain Shah, Vasilis Peristeras, and Ioannis Magnisalis. 2019. Government (Big) data ecosystems: A systematic literature review. In *Proceedings of the International Conference on Digital Information Management*. 1–14.

[93] Ray Wang, Thomas F. Edgar, Michael Baldea, Mark Nixon, Willy Wojsznis, and Ricardo Dunia. 2018. A geometric method regarding batch data visualization, process monitoring and fault detection. *J. Process Contr.* 67, (2018), 197–205. DOI: https://doi.org/10.1016/j.jprocont.2017.05.011

[94] Nishigandha V. Wankhade and Madhuri A. Patey. 2014. Transfer learning approach for learning of unstructured data from structured data in medical domain. In *Proceedings of the 2013 2nd International Conference on Information Management in the Knowledge Economky (IMKE'13)*. 86–91.

[95] Boyi Xu, Li Da Xu, Hongming Cai, Cheng Xie, Jingyuan Hu, and Fenglin Bu. 2014. Ubiquitous data accessing method in iot-based information system for emergency medical services. *IEEE Trans. Ind. Inf.* 10, 2 (2014), 1578–1586. DOI: https://doi.org/10.1109/TII.2014.2306382

[96] H. S. Yang, L. Zheng, and Y. Huang. 2012. Critical success factors for MES implementation in china. In *Proceedings of the IEEE International Conference Industrial Engineering and Engineering Management*. 558–562. DOI: https://doi.org/10.1109/IEEM.2012.6837801⟨/bib>

[97] Jose Jacobo Zubcoff, Llorenç Vaquer, Jose Norberto Mazón, Francisco Maciá, Irene Garrigós, Andrés Fuster, and Jose Vicente Carcel. 2016. The university as an open data ecosystem. *Int. J. Des. Nat. Ecodynam.* 11, 3 (2016), 250–257. DOI: https://doi.org/10.2495/DNE-V11-N3-250-257

[98] Anneke Zuiderwijk, Marijn Janssen, and Chris Davis. 2014. Innovation with open data: Essential elements of open data ecosystems. *Inf. Polity* 19, 1–2 (2014), 17–33. DOI: https://doi.org/10.3233/IP-140329

[99] Anneke Zuiderwijk, Marijn Janssen, Geerten Van De Kaa, and Kostas Poulis. 2016. The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. *Inf. Polity* 21, 3 (2016), 223–236. DOI: https://doi.org/10.3233/IP-160391


Received February 2020; revised September 2020; accepted September 2020