Understanding the development of public data ecosystems

From a conceptual model to a six-generation model of the evolution of public data ecosystems

Lnenicka, Martin; Nikiforova, Anastasija; Luterek, Mariusz; Milic, Petar; Rudmark, Daniel; Neumaier, Sebastian; Kević, Karlo; Zuiderwijk, Anneke; Rodríguez Bolívar, Manuel Pedro

DOI
10.1016/j.tele.2024.102190

Publication date
2024

Document Version
Final published version

Published in
Telematics and Informatics

Telematics and Informatics 94 (2024) 102190

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Understanding the development of public data ecosystems: From a conceptual model to a six-generation model of the evolution of public data ecosystems

Martin Lnenicka [a,b,*], Anastasija Nikiforova [c], Mariusz Luterek [d], Petar Milic [e], Daniel Rudmark [f,g], Sebastian Neumaier [h], Karlo Kević [i], Anneke Zuiderwijk [j], Manuel Pedro Rodríguez Bolívar [k]

[a] University of Hradec Kralove, Faculty of Informatics and Management, Hradec Kralove, Czech Republic
[b] University of Pardubice, Faculty of Economics and Administration, Pardubice, Czech Republic
[c] University of Tartu, Faculty of Science and Technology, Tartu, Estonia
[d] University of Warsaw, Faculty of Journalism, Information and Book Studies, Warsaw, Poland
[e] University of Pristina - Kosovska Mitrovica, Faculty of Technical Sciences, Kosovska Mitrovica, Serbia
[f] Swedish National Road and Transport Research Institute, Gothenburg, Sweden
[g] University of Gothenburg, Swedish Center for Digital Innovation, Gothenburg, Sweden
[h] St. Pölten University of Applied Sciences, St. Pölten, Austria
[i] University of Zagreb Faculty of Geodesy, Zagreb, Croatia
[j] Delft University of Technology, Delft, the Netherlands
[k] University of Granada, Faculty of Business Studies, Department of Accounting and Finance, Granada, Spain

ARTICLE INFO

Keywords:
Public data ecosystem
Evolution
Generation
Conceptual model
Typology
Systematic literature review

ABSTRACT

There is a lack of understanding of the elements that constitute different types of value-adding public data ecosystems and how these elements form and shape the development of these ecosystems over time, which can lead to misguided efforts to develop future public data ecosystems. The aim of the study is twofold: (1) to explore how public data ecosystems have developed over time and (2) to identify the value-adding elements and formative characteristics of public data ecosystems. Using an exploratory retrospective analysis and a deductive approach, we systematically review 148 studies published between 1994 and 2023. Based on the results, this study presents a typology of public data ecosystems and develops a conceptual model of elements and formative characteristics that contribute most to value-adding public data ecosystems. Moreover, this study develops a conceptual model of the evolutionary generation of public data ecosystems represented by six generations that differ in terms of (a) components and relationships, (b) stakeholders, (c) actors and their roles, (d) data types, (e) processes and activities, and (f) data lifecycle phases. Finally, three avenues for a future research agenda are proposed. This study is relevant for practitioners suggesting what elements of public data ecosystems have the most potential to generate value and should thus be part of public data ecosystems. As a scientific contribution, this study integrates conceptual knowledge about the elements of public data

* Corresponding author.
E-mail addresses: martin.lnenicka@uhk.cz (M. Lnenicka), anastasija.nikiforova@ut.ee (A. Nikiforova), m.luterek@uw.edu.pl (M. Luterek), petar.milic@pr.ac.rs (P. Milic), daniel.rudmark@vti.se (D. Rudmark), sebastian.neumaier@fhstp.ac.at (S. Neumaier), karlo.kevic@geof.unizg.hr (K. Kević), A.M.G.Zuiderwijk-vanEijk@tudelft.nl (A. Zuiderwijk), manuelp@ugr.es (M.P. Rodríguez Bolívar).

https://doi.org/10.1016/j.tele.2024.102190
Received 20 November 2023; Received in revised form 23 July 2024; Accepted 19 September 2024
Available online 25 September 2024

ecosystems, the evolution of these ecosystems, defines a future research agenda, and thereby moves towards defining public data ecosystems of the new generation.

1. Introduction

The practice of governments becoming more digital evolves into a new governance paradigm in which a large range of stakeholders engages in public–private collaborations (Clement et al., 2022; Janowski, 2015; Lis and Otto, 2021). In this new paradigm, digital transformation creates an ecosystem where data about all aspects of our world are distributed across multiple information systems (Curry and Ojo, 2020). In data ecosystems, data is the central resource, where access to data can be fully open or more restricted between stakeholders. The ecosystem environment includes the deployment and management of infrastructure resources, as well as activities and tools for interactions between stakeholders and other internal and external elements (Geisler et al., 2022; Heimstädt et al., 2014a; Oliveira and Lóscio, 2018; Van Schalkwyk et al., 2016). When functioning, i.e., reaching a state of equilibrium, data ecosystems enable the creation of innovative products and services, often using standard formats and protocols (Gama and Lóscio, 2014; Oliveira et al., 2019).

Public data ecosystems differ from each other in terms of the composition and importance of elements, the relationships between those elements, and their contribution towards value creation. Exploring these co-evolutionary elements and formative characteristics of data ecosystems is critical to their development and growth in terms of value creation (Azkan et al., 2020). Moreover, existing research on data ecosystems has been noted to be fragmented, both thematically and methodologically (Martin et al., 2017), and is mainly focused on data ecosystems that are open (Susha et al., 2023). A systematic review of the literature on public data ecosystems could provide new insights into their types, formative characteristics, and their resilience over time, including ecosystems with different forms of access to data. Existing efforts to understand how data ecosystems emerge and evolve include Heimstädt et al. (2014a) and Heimstädt et al. (2014b) focused on the evolution of UK open data ecosystems, while Styrin et al. (2017) explored the conditions for ecosystem creation and development. A more recent study was conducted by Gelhaar and Otto (2020), who focused on challenges at the emergence stage, postulating that it is crucial to build trust among ecosystem participants. Oliveira et al. (2019), in turn, conducted a systematic review of data ecosystems, recognizing the importance of theory, models, and engineering, the consideration of which can contribute to a better understanding of how to develop and advance the field of data ecosystem, and concluding that data ecosystem theory is not well developed.

While some studies have advanced an understanding of data ecosystems, there is a lack of synthesis of existing data ecosystem research, emphasising the temporal aspect of ecosystem management, not to say about open and public data ecosystems and their evolution, or its contribution towards value for the data economy (Zillner et al., 2021). As a result, synthesising finding, reconciling conflicting evidence, and drawing a comprehensive understanding of phenomena are crucial for both academic and practitioner communities, providing a strong platform for future research in this area (Palmatier et al., 2018). However, a disconnect in research often leads to redundant investigations, hindering knowledge advancement and leaving certain areas underrepresented, particularly pronounced in the diverse landscape of public data ecosystems. Scattered studies make it challenging to consolidate insights for a particular type of data ecosystem across the body of research, exacerbating this challenge due to various perspectives within the domain. Given the dynamic nature of knowledge generation, systematic reviews become imperative to prevent replicative research that do not substantially advance knowledge, highlighting the research gaps and clarifying directions for future research to substantially advance knowledge (Moher et al., 2009; Paul and Criado, 2020; Paul et al., 2021; Paul et al., 2023).

To fill this gap, this study, first, synthesises the literature, classifying it by the type of public data ecosystem, which also includes definitions of these often similar, but at the same time very different types of public data ecosystems. In this way, the reviewed papers are grouped in a meaningful way to guide the reader toward a better understanding of the phenomenon and provide a foundation for insights about future research directions (Palmatier et al., 2018; Snyder, 2019). Conceptualising a public data ecosystem through the synthesis of multiple perspectives offers clarity and a comprehensive checklist of constituent elements, facilitating deeper understanding and ensuring the inclusion of all pertinent factors. Lastly, the dynamic nature of data ecosystems requires attention yet to be received in the literature, impacting the elements that form these ecosystems and their interrelationships, requires understanding of their development over time.

This study defines and answers three Research Questions (RQ):

RQ1: How are public data ecosystems conceptualised in the literature?
RQ2: What are public data ecosystems' value-adding elements and their mutual relationships?
RQ3: What are the types and evolutionary generations of public data ecosystems over the past 30 years?

To answer the research questions, we use a Systematic Literature Review (SLR) and a deductive approach examining studies published between 1994 and 2023 (inclusively), including an exploratory retrospective analysis of the results. Hence, the study synthesises the literature on public data ecosystems, their formative characteristics and evolution over almost 30 years. Framework-based systematic review (Palmatier et al., 2018; Paul et al., 2023) is conducted having the potential to help researchers, policymakers, and professionals keep track of research findings (Paul et al., 2021) and significantly advance the field being based on a rigorous, transparent, and robust synthesis of past studies and enabling for the development of research agendas that add value.

The findings of this study are relevant for practitioners in developing, designing, and managing a public data ecosystem that has the potential to create value associated with public data ecosystems. Firstly, this study provides practitioners with a better understanding of what the value-adding elements of public data ecosystems are, how these elements are interconnected, and to determine to which evolutionary generation a specific public data ecosystem belongs. Once practitioners have gained these insights, this allows them to identify what value-adding elements are missing or need to be redesigned, as well as what relationships exist between the above elements. These findings are relevant for building and developing the ecosystem as well as redesigning an already existing public data ecosystem.

Ultimately, the findings of this study contribute to the current theoretical body of the knowledge by summarising, structuring, and conceptualising current research in the area of public data ecosystems. As a scientific contribution, this study defines the evolution of public data ecosystems, their elements, the relationships between them, and the milestones that trigger the transition from one generation of the ecosystem – more static in the past – to another, including the recent and forward-looking generation that is more dynamic and that focuses on modern technologies, stakeholders' engagement, and value (co–)creation. Finally, this study identifies a future research agenda to improve the practical understanding and application of public data ecosystems by addressing real-world challenges and contributing to the practical application of scientific findings, especially within the forward-looking generation of public data ecosystems.

The remainder of the paper is structured as follows: Section 2 defines the key concepts to establish their common understanding, Section 3 outlines the research approach, describing the concrete steps. Section 4 presents the results of the SLR and their analysis. Section 5 presents discussion, future research agenda, limitations, and theoretical and practical implications of the study, while Section 6 concludes the paper.

2. Research background

This section provides the background for this study. Before we discuss how the literature defines public data ecosystems, we define the general concepts of data ecosystems and data infrastructures.

2.1. Data ecosystem

The data ecosystem, of which public data ecosystems are subsets, has been a constant topic of interest in many areas, with various definitions emerging over the years. In the context of the information network of relationships, an ecosystem can be defined as "*a system of people practices, values and technologies in a particular local environment*" (Nardi and O'Day, 1999, p. 49), wherein the present case, the environment of the ecosystem is government, at all levels, including its organisational practices, policies, and technical platforms developed to facilitate the development of this ecosystem (Dawes et al., 2016). From a more technical point of view, data ecosystems are virtual data spaces utilising current technology and standards, along with recognized governance frameworks for the data economy, to enable safe, standardised data interchange and simple data linking (Petersen et al., 2019). Furthermore, data ecosystems are distributed, open, and adaptive information systems that possess self-organising, scalable, and sustainable characteristics (Geisler et al., 2022).

In comparison to "*business ecosystems*" and "*software ecosystems*", data ecosystems depend on a large and diverse group of actors, each with unique traits, skills, and expectations. Data ecosystems deal with heterogeneous resources, where actors may use a variety of methods and environments to create and consume resources, as well as in different settings, where most elements are dynamic and evolve over time (Oliveira and Lóscio, 2018). Data ecosystems encompass distributed, heterogeneous, dynamic, and evolving actors and resources constituting "*a set of networks composed by autonomous actors that directly or indirectly consume, produce, or provide data and other related resources (*e.g.*, software, services, and infrastructure). Each actor performs one or more roles and is connected to other actors through relationships, in such a way that actors' collaboration and competition promotes data ecosystem self-regulation*" (Oliveira and Lóscio, 2018, p.4). With such an approach, the main elements of the data ecosystem are actors, roles, relationships, and resources, so their definitions are nested within the actor–network theory and consider the social and technical relationships between elements, how they interact and collaborate with data resources throughout their lifecycle to create value (Oliveira et al., 2018; Oliveira et al., 2019).

Kontokosta (2013) claims that the way we use the data ecosystem can alter market behaviour, and individual decision models and contribute to public gain by supporting the decision-making processes of each end user and influencing both the producers and consumers of these data. From the institutional point of view, a data ecosystem can be used as a means for supporting decision-making and planning (Haak et al., 2018). Similarly, Gelhaar and Otto (2020) point out that data ecosystems are derived from business ecosystems that aim to develop new value propositions that serve as the foundation for future innovation by collaborating with both supply and demand parties. Moreover, they argue that it is crucial to build trust among ecosystem participants.

The essential components of the data ecosystem are the relationships between data users, data providers, tools, and the data infrastructure (see the next section) (Charalabidis et al., 2018; Diran et al., 2020). Curry and Sheth (2018) believe that massive collaboration within a data ecosystem should be accompanied by a proper data governance model that fully considers ethical, legal, and privacy concerns. McLoughlin et al. (2019) go even further, stating that data ecosystems can be understood as interpretive communities that give meaning to data and how they should be used, linked, shared, and – according to Wilson (2019) – how to produce meaningful insights from data. Relationships facilitate the creation, processing, and use of data between actors for analytical purposes that create added value for all stakeholders (Azkan et al., 2020; D'Hauwers et al. 2022). Those complex interconnections between the actors within the ecosystem lead to situations in which actors are working cooperatively and competitively at the same time – also known as *coopetition* (Gelhaar et al., 2021).

2.2. Data infrastructure

As several definitions presented in the previous sections demonstrated, some studies tend to define the data ecosystem through or using the term "*infrastructure*", which is considered to be a foundation necessary for the operation of society or enterprise, and which, together with data ecosystems, is claimed to be viewed as a metaphor deployed in a twin manner (Davies, 2011). Infrastructures can be of several types, where in the context of data ecosystems, with two of the most important in the context of data ecosystems being *data infrastructure* and *digital infrastructur*e, where data infrastructure is a public good, which is part of the digital infrastructure. The concept of *national data infrastructure* is more open in terms of data, implementation alternatives, application sectors, and aims (Klievink et al., 2017). It is a distributed technical infrastructure (comprising portals, platforms, and services) that enables data access and exchange depending on predefined rules (Estermann et al., 2018). According to Gulson and Sellar (2019), data infrastructures establish novel relationships between various stakeholders and generate forces that both produce and operate across these network spaces.

Gelhaar et al. (2021) claim that the data infrastructure is provided by a platform that supports the sharing and use of data within an ecosystem. Platform, in turn, is defined as a coherent / holistic part of the ecosystem in terms of its features and capabilities for working with and reuse open datasets by stakeholders, which is usually technological in nature (Bonina and Eaton, 2020; Danneels et al., 2017). Nevertheless, infrastructure is an important prerequisite for building and enabling data ecosystems and represents a step forward in the design of intelligent systems, i.e., smart systems.

2.3. Public data ecosystem

A public data ecosystem represents a distinct subset of a data ecosystem, primarily composed of data sourced from and/or funded by governmental entities (central and local government or any other public body), which are also responsible for setting the legal framework governing its operation. Since public data ecosystems rely heavily on currently available technology, there are gaps in the research body in this field arising from a lack of common understanding of the public data ecosystem, particularly one that should be compliant with current trends, and its elements. For example, there is very limited research on how conceptual approaches to Open Government Data (OGD) governance should incorporate Artificial Intelligence (AI)-related concepts, especially in the way they change the data acquisition and data processing (Tan, 2022).

Thus, the public data ecosystem is understood by us as *a dynamic and adaptable network of elements and interrelations between them, driven by occurring internal and external data flows and requirements arising within it. Element*, in this definition stands for each item / component that can be used to describe the context of an ecosystem and / or to affect (influence) the dynamics of data ecosystem management and development. Interrelation between these elements or relationships is a connection between elements, the nature of which depends on the nature of the elements with which it is connected. Most of the relationships are represented by data activities through which data are disclosed and reused (McLeod and McNaughton, 2016). Dynamism and adaptability of the network of these elements and interrelations between them, in turn, refers to changes in the state of a data ecosystem that occur through a series of successive / sequential steps over time, including its key milestones, which is called *evolution*.[1]

3. Research approach: Systematic literature review

SLR approach was used to answer or build the foundation for answering our RQs. More precisely, objectives of our literature review were to (1) understand how are public data ecosystems conceptualised in the literature studying all relevant literature covering this topic (RQ1), (2) identify value-adding elements that constitute public data ecosystems and mutual relationships between them (RQ2), (3) identify how public ecosystem evolution is defined in the literature, what are the types of public data ecosystems, and what are the formative characteristics affecting or triggering the transition from one evolutionary generation to another (as part of RQ3). Following Kitchenham (2004), the SLR is conducted in five steps: (1) study identification, (2) study selection, (3) study relevance and quality assessment, (4) data extraction and (5) data synthesis, which are documented in detail in Supplementary Materials (Appendix A).

To identify relevant literature, the SLR was carried out to form the knowledge base by querying digital libraries covered by Scopus and Web of Science (WoS). Given the specificity of the topic, we searched the Digital Government Research Library (DGRL). Finally, we searched Google Scholar.

The search query (see Supplementary Materials, Table A.1 (Appendix A)) was defined as a combination of terms "*public*" and "*data ecosystem*", where the latter has been provided with various alternative namings found in the literature, e.g., Oliveira et al. (2019), and based on our own experience, including "*data infrastructure*",[2] "*data space*", "*data collection ecosystem*", "*dataset ecosystem*", "*data on the web ecosystem*". In the case one of the above terms referred to a topic other than the public data ecosystem, this study was excluded from further analysis after the first round of results filtering. As such, the "("data ecosystem" OR "data infrastructure" OR "data space" OR "data system" OR "data collection ecosystem" OR "dataset* ecosystem" OR "data set* ecosystem" OR "data on the web ecosystem")

[1] definition adapted from the Cambridge Dictionary definition of "evolution" defined as (1) "*a gradual process of change and development*", 2) "*evolution is the process by which the physical characteristics of types of creatures change over time, new types of creatures develop, and others disappear*", https://dictionary.cambridge.org/dictionary/english/evolution

[2] Although "*data ecosystem*" and "*data infrastructure*" cannot be considered synonymous given their semantic meaning, for the sake of completeness of the results, we included these terms since researchers sometimes use them interchangeably.

AND ("public")" query was used. The SLR was first conducted in September 2022, and then, considering the dynamism of the topic and the ongoing research in this domain, we updated the search in September 2023, to build a comprehensive knowledge base.

After selecting studies, assessing their relevance and quality, resulting in 148 studies, and extracting data using a protocol we designed (see Supplementary Materials, Table A.2 (Appendix A)), we systematically analysed the obtained raw data (see Section 4). Then, by synthesising the literature review, we designed conceptual models of public data ecosystems and a conceptual model of the evolutionary generations of public data ecosystems. These models are designed using the deductive approach, which entails "*identify [ing]dimensions and characteristics […] by a logical process derived from a sound conceptual or theoretical foundation*" (Nickerson et al., 2013, p. 340).

The procedure used to design it consisted of several activities. First, each of the authors of the study individually reviewed the information collected about each of the selected papers. By comparing the data derived for each metadata dimension (see Supplementary Materials, Table A.2 (Appendix A)) for 148 studies, each author derived patterns. These were discussed, and prioritised among the first two authors, and then one author took the lead in developing the draft model. Then, another author was asked to review the model. Then, eight authors were asked to review the model, its elements, the relationships between them, and the milestones that triggered transition from one generation of the ecosystem to another, and provide the feedback on it, suggesting changes – adding new elements, relationships, or milestones, suggesting removing one of the above, or modifying. Then, all suggestions were incorporated in the revised version of the model, the two authors reviewed the model, and the model has been reviewed by the authors again, leading to the final version of the model. As such, the design of the model is based on the informed assessment of 148 studies and further analysis with the authors of this study.

4. Literature review – Results and their analysis

This section outlines the results we obtained and conclusions we reached after analysing the selected studies. As part of the study, we conducted bibliographic analysis, descriptive analysis, approach- and research design- related information analysis, and public data ecosystem-related information analysis. The results of the former parts are available in the Supplementary Materials (Appendix B), while here we focus on the RQs-related aspects of our SLR, namely, the analysis of public data ecosystem-related information, reflecting only briefly on the former parts. The data underlying our study are publicly available through Zenodo.[3]

The popularity of the topic of public data ecosystems has grown since the early 2010s, with further significant increase in recent years (starting in 2019), with the number doubling and tripling since the 2010s. Examination of the objectives and contributions made by 148 selected studies with heir further classification by topic they focused on, coding these studies in several iterations (Supplementary Materials (Appendix B, Table B.1)), made it apparent that existing studies cover a wide range of categories (we classified them into 15 categories) with varying degrees of popularity (from 3 to 19 studies in each category). Among them, some topics seem particularly important and popular, while others promise future research and application. At the forefront of scholarly attention in terms of the number of studies addressing relevant topics are studies focusing on the **development and evaluation of frameworks** tailored to understanding, analysing, and managing data ecosystems in use (19 studies). The popularity of this topic underscores the importance of establishing robust frameworks for navigating the complexities inherent in data ecosystems. This emphasis on theoretical foundations is complemented by **empirical examinations and case studies** that provide real-world insight into the functioning, challenges, and opportunities in data ecosystems in often very specific contexts or domains. These studies offer practical insights that are critical for effective decision-making and policy formulation, although they often remain relevant to the context being studied. In addition, analyses focusing on **policy formulation and implementation** have received attention as understanding the impact and effectiveness of policies governing data ecosystems is imperative to ensure responsible and effective management of the entire data ecosystem, data sources, or other specific components of these ecosystems. In addition, healthcare, social services, and environmental sustainability may gain greater popularity as one of the key types of public data ecosystems, as well as the ongoing evolution of technology necessitates deeper exploration into the technical components and system development within data ecosystems.

Primarily, the studies conducted qualitative research ($n = 125$, which is 84.5 %), which might be due to the explorative nature of public data ecosystem research. Other studies leveraged mixed research methods or quantitative research. Literature reviews, content analysis, surveys, and interviews emerged as the predominant methods. Upon examination of whether the selected studies referred to any theory and/or theoretical concepts and/or approaches, we observed that almost half of the studies ($n = 63$, which is 42.6 %) omitted reference to these. Of the remainder, the ecosystem approach, ecosystem thinking approach, systems theory or platform theory were most frequently employed.

Most studies predominantly employ a country-level focus, typically describing and/or comparing at least two countries (use cases). The second most common case envelops various enterprises within a particular industry or individual organisations within specified public sector domains.

4.1. *Research question 1: Conceptualization of public data ecosystems*

This section answers our first research question: *How are public data ecosystems conceptualised in the literature?* During our analysis of

[3] https://zenodo.org/doi/10.5281/zenodo.13842001

selected studies, we identified various conceptualizations of public data ecosystems in the scientific literature that we classified into categories depending on their focus, unique characteristics, stakeholders, governance mechanisms, and challenges by which they are characterised in the respective studies (the full version of the table with all relevant studies by category is available in Supplementary Materials (Appendix C, Table C.1)). Our analysis revealed a diverse landscape of data ecosystem types. Among them, some types have garnered significant attention in the form of studies, while others, even if not as popular, offer promising opportunities for future research and application. At the forefront of research interest are Spatial Data Infrastructure (SDI) and OGD ecosystems.

The oldest type that we identified is the **geodata / spatial / geospatial data ecosystem**, which since the early 1990 s has been implemented as the SDI connecting geodata / spatial / geospatial data, Geographic Information Systems (GIS), users and tools. SDI can be defined as a set of networks and infrastructures that link and integrate vertical (local, regional, national) and horizontal (data categories) (McLaughlin and Nichols, 1994) components. This definition was later extended to encompass the ecosystem approaches and corresponding processes such the facilitation and coordination of data exchange and sharing between stakeholders (Mulder et al., 2020; Rajabifard et al., 2002; Shakeri et al., 2013; Vancauwenberghe et al., 2014), concepts such as open government, open data, and open-source software (Sveen, 2017), and respect the needs and demands of users to collect, manage, distribute/disseminate, and more effectively utilise geodata and associated services (Crompvoets et al., 2011; Grus et al., 2010; Kok and Van Loenen, 2005). Moreover, it expanded to include various policies, frameworks, technologies, systems and infrastructures, the financial and human resources required to guarantee that stakeholders working with these data, whether at a local, regional, national, or global level, are not limited in achieving their goals (Kalliola et al., 2019; Parida and Tripathi, 2018). Overall, these ecosystems play a pivotal role in supporting applications ranging from urban planning to environmental management.

As open government and transparency initiatives have evolved around the world, a new type of public data ecosystem has emerged – **OGD ecosystem**. However, we identified that over time, three distinct ecosystems have materialised in the literature. The first was the **open government ecosystem**, in which data was not the most important element, but the openness of this type of ecosystem was ensured by supporting the flows of information and data. Harrison et al. (2012) argued that the interactions between governments and other public officials with citizens, businesses, and civil sector organisations drive the dynamics of these flows. After establishing open data principles, two other types have come to the forefront, i.e., **open data ecosystem** and **OGD ecosystem**. The main difference between the two is that while the OGD ecosystem is usually strictly focused on data provided by governments, either national, regional, or local, open data ecosystems include more data meeting open data principles that are produced and disclosed by other stakeholders such as research institutions, and businesses. With a significant number of studies dedicated to this type, OGD ecosystems are recognized as crucial for fostering transparency, innovation, and improved public services.

Additionally, **Big Data ecosystems** have received attention, reflecting the growing importance of big data in contemporary discourse. These ecosystems encompass the collection, analysis, and use of large volumes of data to extract insights and support decision-making across various domains. In other words, they can be described as complex multi-layered, data-intensive digital-physical systems (Orenga-Roglá and Chalmeta, 2019; Shah et al., 2021b) that comprises a set of interdependent components used throughout the data lifecycle to cope with the evolution of data from multiple data sources, models, applications, services, and associated infrastructures (Demchenko et al., 2014; Orenga-Roglá and Chalmeta, 2019; Shah et al., 2021a), which desired characteristics are (1) robustness and fault tolerance, (2) low-latency reads and updates, (3) scalability, (4) generalisation, (5) extensibility, (6) ad hoc queries, (7) minimal maintenance, and (8) debuggability (Singh et al., 2019). In recent years, research attention has shifted towards the delivery of outputs and their utilisation for business intelligence and decision support systems (Munshi, 2018; Pinto and Parreiras, 2020; Yunita et al., 2022). Likewise, **Linked Open Data (LOD) ecosystems** as a subset of the OGD ecosystem increases in popularity, recognizing the importance of data interoperability across diverse sources and domains (Ding et al., 2012; Song and Kim, 2013).

Looking to the future, several types of **domain-specific data ecosystems** offer promising avenues for future exploration. I**nnovation data ecosystems** that focus on collecting and analysing data related to innovation processes have the potential to drive economic and social progress (Jetzek, 2017). Another types are **health(care) data ecosystems** (Aaen et al., 2022; Hoeyer, 2020; Park and Gil-Garcia, 2017), **environmental data ecosystems** (Kontokosta, 2013; Lienert and Meier, 2014), **educational data ecosystems** (Gulson and Sellar, 2019; Hardy and Liu, 2022; Shah et al., 2021a; Van Schalkwyk et al., 2016), **economic data ecosystems** (Zuiderwijk et al., 2016), **tourism data ecosystems** (McLeod and McNaughton, 2016), and **chemical engineering data ecosystems** (Ramalli and Pernici, 2022). Environmental data ecosystems offer opportunities to address pressing environmental sustainability issues. By facilitating informed decision-making and conservation efforts, these ecosystems can help mitigate environmental challenges. Health(care) data ecosystems are poised for further exploration given the critical role of healthcare data in healthcare delivery. Research in this area can contribute to advances in personalised patient care and health policymaking. Similarly, educational data ecosystems can help improve teaching, learning, and education outcomes through data-driven approaches. By leveraging data, these ecosystems can inform personalised instruction, curriculum development, and educational policy-making.

Examples of **smart city and Internet of Things (IoT) data ecosystems** may gain popularity in the context of the current trend toward the smartization of cities and the increase in the number of connected devices producing data. Thus, both a combination of the concepts of IoT and Smart City, as well as two subtypes of (1) **IoT and open data ecosystem** and (2) **smart city data ecosystem** may be expected as gaining popularity in the future, where IoT and Open Data ecosystem involves the integration of IoT devices with open data platforms, enabling the collection, analysis, and sharing of real-time data to improve decision-making, services, and efficiency in various domains (e.g., environmental monitoring) (Kalaitzakis et al., 2019; Curry and Ojo, 2020), while Smart City Data Ecosystem involves the collection, integration, and analysis of data from a variety of sources within urban environments to improve infrastructure, services, and quality of life for residents (Gupta et al., 2020).

Finally, **fair data ecosystems**[4] that emphasise fairness and transparency in data governance are likely to gain traction amid growing concerns about data ethics and privacy as an alternative to the citizen-centric data ecosystem and digital single market that the European Union seeks to establish to enable the creation of public value and empowerment of citizens (Calzati and van Loenen, 2023). This fair data ecosystem is rooted in the data republic model and is suggested to be built around data commons coupled with open data frameworks and SDIs. Fair data ecosystem includes roles, rules, and mechanisms to systematically consider and, if necessary, arbitrate between the interests of the data and the values of the involved actors, while maintaining a collective outlook (Calzati and van Loenen, 2023). As such, research in this area can contribute to fairer and more responsible data practices. In addition, recently, a new type of the public data ecosystem has emerged termed the **(public) data space** (Tron, 2020), which is defined as large digital ecosystems that are focused not only on the economic exchange but also other network effects resulting from cooperation between groups of actors (Beverungen et al., 2022).

Although several types of data ecosystems were identified, there remains many opportunities for research and innovation in a variety of areas. By delving into both existing and new types of ecosystems, researchers can contribute to advancing knowledge and realising the transformative potential of public data ecosystems. To this end, based on the classification of studies by the type of data ecosystem they cover (conceptualising understanding of its types) and using a deductive approach, we define a typology of public data ecosystems, presented in Fig. 1. These types include the geodata / spatial / geospatial data ecosystem, considered the oldest public data ecosystem, and the open government ecosystem, which includes the open data ecosystem, the OGD ecosystem, which can be extended by the geospatial data ecosystem, and Linked Open Government Data (LOGD) ecosystem as an extension of the previous one. In addition, the big data ecosystem can extend the LOGD ecosystem, and can be extended by the Smart City and IoT data ecosystem. A set of ecosystems of different types, namely, big data ecosystem, domain-specific data ecosystem, stakeholder-centred data ecosystems / collaborative data partnerships, Smart City and IoT data ecosystem, and policy and governance data ecosystems, in turn, can extend another type of data ecosystems – the infrastructure data ecosystem. Less mature but very promising public data spaces and the fair data ecosystem are also included, where the latter can extend the geospatial data ecosystem and policy and governance data ecosystem. These are 13 types of public data ecosystems, all of which contribute to expanding the boundaries of public data ecosystems, being their subtypes.

4.2. Research question 2: Value-adding elements of public data ecosystems and their mutual relationships

This section answers our second research question: *What are public data ecosystems' value-adding elements and their mutual relationships?*

First, we briefly summarize the results of the SLR (with extended version reflecting on the most relevant studies available in Supplementary Materials (Appendix D)), on which basis we then present a conceptual model of value-adding elements of public data ecosystems and their mutual relationships.

4.2.1. Value-adding elements of public data ecosystems and their mutual relationships in the literature

Public data ecosystems are complex phenomena influenced by a multitude of factors across various dimensions. Firstly, the establishment and success of these ecosystems depend heavily on administrative capacities, especially in IT-related areas, and government initiatives such as e-government (Ahn and Chu, 2021; Estermann et al., 2018). Additionally, the power balance among data subjects, controllers, and processors plays a crucial role, necessitating alignment with quality reports, cross-sectoral monitoring, and coherent service administration (Aaen et al., 2022). Proper management of interconnected components, stakeholders, and data dynamics is essential to ensure privacy, transparency, and trust while harnessing data for public benefit (Liva et al., 2023).

Furthermore, **stakeholders** in public data ecosystems, including governments, academia, civil society organizations, businesses, and citizens, play diverse roles spanning data provisioning, consumption, aggregation, and sponsorship (Jetzek, 2017). Developers are particularly crucial in leveraging government data to create innovative solutions (Jean-Quartier et al., 2022). Academia contributes to research, analysis, and policy recommendations, acting as a bridge between technical communities and policymakers (Shah et al., 2021b). **Actors** within the ecosystem, identified by various researchers, include data providers, consumers, users, public managers, businesses, civil society, and citizens, each playing distinct roles in shaping data needs, transparency preferences, and innovation (Harrison et al., 2012; Reggi and Dawes, 2022; Schneider and Comandé, 2022).

Regarding **data**, studies discuss various types such as spatial, statistical, health, environmental, transportation, economic, big data, linked data, IoT, and real-time data, highlighting the importance of data openness and technological advancements such as AI, IoT, and cloud computing in handling and analyzing diverse data sources (Gelhaar et al., 2021; Van Loenen et al., 2021). **Processes and activities** within the ecosystem, such as data collection, sharing, cooperation, and innovation, are critical for value creation and sustainability (Xiao and Miller, 2021; McBride et al., 2020). In addition, data ecosystems can evolve based on the specific **data lifecycle**

[4] Findability, Accessibility, Interoperability, Reusability (FAIR) of data – set of principles introduced by Wilkinson et al. (2016) to improve the infrastructure supporting the reuse of scientific / scholarly data, including enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals, with further definition of FAIRness governance in (Wilkinson et al., 2022).

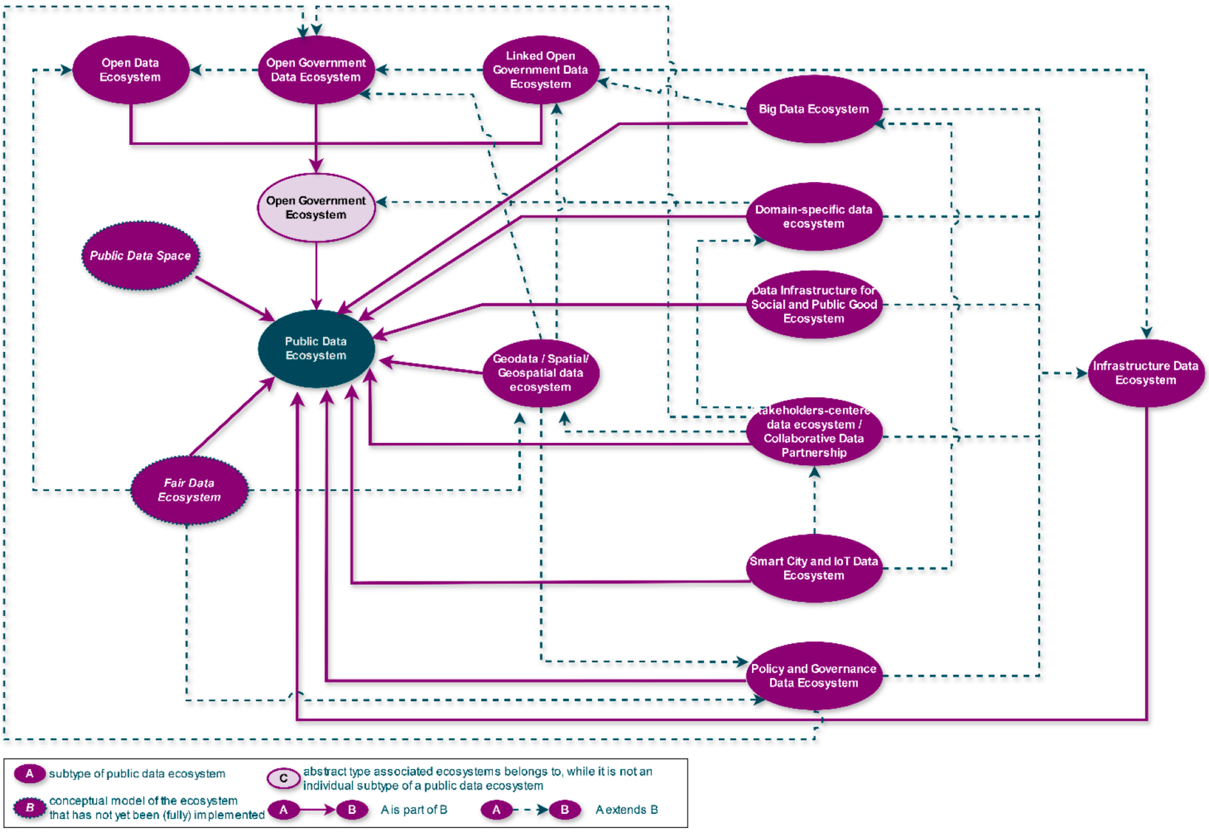

Fig. 1. Typology of public data ecosystems.

phases they focus on, which may include *data request*, *retrieval*, *filtering*, *acquisition*, *extraction*, *management*, *preparation*, *storing*, *archiving*, *processing*, *analysis*, *visualisation*, *linking*, *publication*, *evaluation*, *discussion*, *sharing*, and *reuse*. These phases determine how data are accessed, processed, and used within the ecosystem. Data ecosystems may also give precedence to certain processes, which then can serve as goals to achieve or become driving forces for further improvement and development. These processes could include *finding*, *archiving*, *publishing*, *consuming*, or *reusing data*, as highlighted by Linåker and Runeson (2021). Hence, prioritising particular processes can sculpt the ecosystem's goals and steer its continued evolution, where transparency, accountability, efficiency, and ethical considerations are essential elements in ensuring the ecosystem's success (Noorman, 2017; Hrustek et al., 2023).

Public data ecosystems operate at various **levels**, including national, city, and international levels, each striving for sustainability, stability, and smartness, with sustainability being a key metric to assess ecosystem health (Tang et al., 2022). The social context within which these ecosystems operate is crucial, considering features such as complexity, adaptability, self-organization, and the interplay between government intentions, value creation, and sustainability (Bonina and Eaton, 2020; Azkan et al., 2020). Finally, ensuring data quality, trust, interoperability, and convenient access through collaboration and stakeholder participation are fundamental for a well-functioning ecosystem (Diran et al., 2020; Wilson and Cong, 2021).

4.2.2. Conceptual model of a public data ecosystem

To summarise our findings, using a deductive approach, we conceptualised the formative characteristics identified in the SLR in the conceptual model of the public data ecosystem presented in Fig. 2. The SLR results show that the public data ecosystem is shaped by different elements, which can be divided into three groups or *meta*-characteristics.

The first group is **data**, which can be detailed into (1a) *data characteristics*, such as data types, data formats, data sources, data volumes, data quality, data topics and data categories, data dynamism, (1b) *data infrastructure*, which covers technical equipment, including network and scalability and architecture and environment, as well as platforms, tools and sources, including their features and capabilities, management, security, guidelines and support, and technical standards encompassing various aspects including data formats, protocols, metadata schemas, and semantic standards that focus on encoding meaning into data, adhering to common

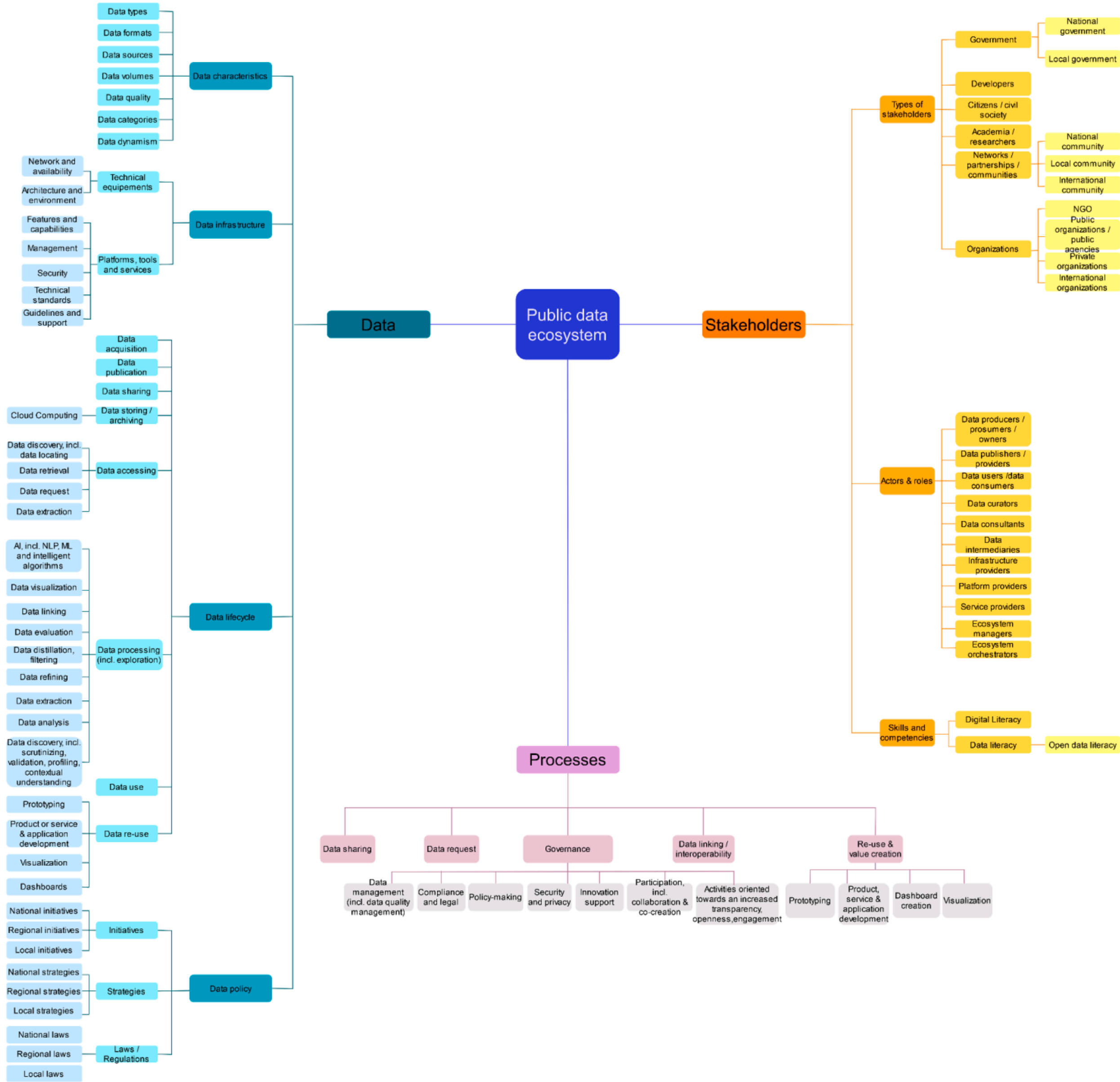

Fig. 2. Conceptual model of a public data ecosystem based on the SLR (for better resolutions see https://zenodo.org/doi/10.5281/zenodo.13842001).

vocabularies, ontologies, and semantic models, allowing for the explicit representation of concepts, relationships, and semantics within dataset, enabling effective data exchange, discovery, and utilisation, (1c) *data lifecycle* that includes data acquisition, data publication, data sharing, data storing / archiving, data accessing, data processing (including exploration), data use and data reuse,[5] and (1d) *data policy*, which includes initiatives, strategies and laws of national, regional and international levels. These categories can be further divided, as illustrated in Fig. 2.

The second group pertains to **stakeholders**. This group encompasses several perspectives, namely, (2a) *types of stakeholders*, which includes government, developers, citizens / civil society, academia / researchers, networks / partnerships / communities, and organisations that includes both NGOs, public organisations / public agencies, private organisations, and international organisations, where government and networks / partnerships / communities can be of local and regional levels, while the latter also of international

[5] **data use** refers to the initial utilisation of data (incl., accessing, analysing, interpreting, and deriving insights from data), while **data reuse** refers to the broader repurposing of data of data beyond its original purpose or context (incl., sharing, integrating, or combining data with other datasets or applications to generate additional value or insights).

level, (2b) *actors and roles*, which are data producers (also prosumers and owners), data publishers / providers, data users (also data consumers), data curators, data consultants, data intermediaries, infrastructure providers, platform providers, service providers, ecosystem managers, and ecosystem orchestrators,[6] (2c) *skills and competencies*, which includes digital literacy and data literacy, which includes an emerging subtype of open data literacy.

Finally, the third group is **processes**. This group encompasses (3a) *governance*, which includes data management (including data quality management), compliance and legal processes, policy-making, security and privacy, innovation support, participation, including collaboration and co-creation, and other activities oriented towards an increased transparency, openness, and engagement, (3b) *data requesting*, (3c) *data sharing*, (3d) *data linking and interoperability*, and (3e) *data reuse and value creation*.

Although some components can be broken down into sub-components, this conceptual model strives to satisfy the qualitative attributes of a model to be deemed "good". Such attributes, also referred to as the necessary conditions for a model, ensure its utility. According to Nickerson et al. (2013, p. 341), a model should be (1) *concise* – have a limited number of dimensions and a limited number of characteristics in each dimension to prevent cognitive overload and future comprehensions or application challenges, (2) *robust* – sufficient number of dimensions and characteristics to clearly distinguish objects of interest, as long as this does not compromise its conciseness (Bailey, 1994, p. 1), (3) *comprehensive or complete* – able to classify all known objects within the domain in question (Bowker, 2000), (4) *extendible* – permitting additional dimensions and new characteristics to be included in the dimension as new types of objects appear, and (5) *explanatory* – contains dimensions and characteristics that do not describe all possible details of the objects, but rather provide useful explanations of the nature of the objects being studied or future objects to help us understand these objects. Thus, we decided to limit the number of levels for identified dimensions – data, stakeholders, process – to a maximum of four levels. This does not mean, however, that this fourth level is the last one that can be defined, since the model can be expanded (in breadth and depth). For example, the technical standards component (part of the data infrastructure subcomponent) can be expanded by adding the above-mentioned subcomponents, namely data formats, protocols, metadata schemas, and semantic standards that, in turn, could lead to the definition of the sixth level that would include common vocabularies, ontologies and semantic models, while the data retrieval component can be broken down into download and API retrieval, etc.

4.3. Research question 3: Types and evolutionary generations of public data ecosystems

This section answers our third research question: *What are the types and evolutionary generations of public data ecosystems over the past 30 years?* First, we present the results of the SLR by reflecting on the selected studies, which then serve as an input for the conceptual model that captures evolutionary generations of public data ecosystems designed using the deductive approach.

4.3.1. Evolution of public data ecosystems in the literature

The evolution of public data ecosystems, their elements and formative characteristics are primarily shaped by the technological dimension and the level of its penetration among stakeholders, specifically, what technologies, tools, standards etc. are available to operate these ecosystems and the degree to which stakeholders can use them effectively and efficiently (Beverungen et al., 2022; Elwood, 2008; Heimstädt et al., 2014b; Klievink et al., 2017; Linåker and Runeson, 2020; Mulder et al., 2020; Rajabifard et al., 2002; Vancauwenberghe and van Loenen, 2018). As information and database systems have progressed over time, they have reshaped how public data are processed, stored, and shared. The emergence of networked data infrastructures, particularly with the advent of the Internet, has played a pivotal role in facilitating data transmission and sharing among stakeholders. I.e., since the 1970s these systems began to be connected through networks, and in the early 1990s, with the rise of the Internet, public data began to be transmitted and shared among stakeholders through these networks. These infrastructures are essential for launching, operating, and interacting with elements reliant on them. Additionally, the 21st-century data revolution, characterized by advancements such as AI, IoT, and cloud computing, continues to shape the evolution of public data ecosystems, influencing how data are acquired, processed, stored, managed, and shared.

Furthermore, beyond technological dimensions, various drivers contribute to the creation and development of public data ecosystems, including digital technology emergence and political and institutional efforts (Kalaitzakis et al., 2019). Government policies, data management practices, and stakeholder involvement are key determinants in shaping these ecosystems. The evolution from spatial data infrastructures (SDIs) to open data ecosystems reflects a transition towards more open and transparent data-sharing practices, aligned with principles of accountability and openness (Welle Donker, 2016). As these ecosystems continue to evolve, concepts like sustainability and resilience come into focus, emphasizing the importance of government intervention and collaborative efforts among stakeholders to ensure the continued development and functionality of these ecosystems (see more detailed description of this evolution in Supplementary Materials (Appendix E)).

As such, we can define the **evolution of public data ecosystems** as *the developments of their elements over time, enabled by arising and emerging technologies and interactions that occur within ecosystems to achieve their purpose*. Considering this aspect, it is easier to understand evolutionary stages and development milestones (Heimstädt et al., 2014a; Heimstädt et al., 2014b; Jean-Quartier et al., 2022). Evolution is based on the involvement of other stakeholders outside the public sector in the ecosystem and its development (Liva et al., 2023; Vancauwenberghe and van Loenen, 2018). The evolution is also known to be externally affected by comparisons of performance and impact, and benchmarking with other countries, as well as best practices (McBride et al., 2020; Styrin et al., 2017).

[6] It's worth noting that in certain ecosystems, some of these roles can overlap or merge.

4.3.2. Conceptual model of the evolutionary generations

As the result of the SLR presented in the previous sections, and considering the ecosystem perspective, the conceptual model was developed to capture evolutionary generations of public data ecosystems that we call Evolutionary Model of Public Data Ecosystems (EMPDE) (Fig. 3). These generations are described based on their value-adding elements and formative characteristics that represent the public data ecosystem in each generation. As with the previous conceptual model, the deductive approach was used to design it. The SLR results show that the evolution of public data ecosystems concerns several value-adding elements and formative characteristics, which can be divided into five groups or *meta*-characteristics that represent the public data ecosystem in each generation. These *meta*-characteristics are: (1) *formative components and relationships*, (2) *stakeholders*, (3) *actors and their roles*, (4) *data types*, (5) *processes and activities and/or data lifecycle phases*. We chose an additive approach, i.e., each successive generation includes characteristics of the previous generation and at the same time adds new characteristics that best describe it. The boundaries between generations are not clearly defined (therefore only a dashed line separates them) and at the same time not every subsequent generation necessarily contains the previous characteristics.

Eight authors of this paper were asked to review the model, its elements, the relationships between them, and the milestones that triggered transition from one generation of the ecosystem to another, and provide the feedback on it, suggesting changes – adding new elements, relationships, or milestones, suggesting removing one of the above, or modifying. In other words, both substruction and reduction (Bailey, 1994, p. 24) were conducted until a satisfactorily complete model was reached. Then, all suggestions were incorporated in the revised version of the model, the two authors reviewed the model, and the model has been reviewed by the authors again, leading to the final version of the model. Finally, each author provided an opinion on the relevance of the model, each generation, components, and relationships between them for actual data ecosystems they are familiar with.

Based on the generalisation of the concepts identified during the SLR, we identified six evolutionary generations (see also Fig. 3):

1. **the first generation** — "**raw data-centred generation**" (typically before the 1990s to 1991/1993) was represented by stand-alone and usually isolated data infrastructures, in which only governments and some public organisations focused on data generation and transfer (in terms of digitization of analogue data), data were used only within internal networks;
2. **the second generation** — "**(geo)spatial data-centred generation**" (typically from 1991/1993 to 2000s) is characterised by the increasing popularity of SDIs. In this period, data policy and other related policies were established to govern the (geo)spatial data.

Characteristics forming the evolution of generations	Evolutionary generations of public data ecosystems					
	1. GENERATION raw data-centred	**2. GENERATION** (geo)spatial data-centred	**3. GENERATION** public data sharing	**4. GENERATION** open (government) data	**5. GENERATION** public data-driven	**6. GENERATION** intelligent public data
Components and relationships	Data resources Data infrastructures +	Data policy and other related policies Data lifecycle management (data governance) +	Technologies (platforms, tools, and services) Technical standards and guidelines +	Data-related competencies and skills Dynamics of processes and activities +	Big and open linked data, cloud computing External pressures – political, economic, environmental, ethical, and legal +	Intelligent algorithms, machine learning, artificial intelligence, and natural language processing tools
Stakeholders	Governments (national) Public organizations +	Governments (regional, local) Citizens Academia +	Private organizations NGOs +	Developers +	International stakeholders +	Networks, partnerships etc.
Actors and their roles	Data producers Data owners Internal data users +	Data providers External data users +	Policies, laws, and rules parties Data publishers +	Ecosystem orchestrators Data prosumers Data intermediaries +	Infrastructure providers Platform providers Service providers +	Ecosystem managers Data curators Data consultants
Data types	Data (no specific type) +	(Geo)spatial data +	Public sector data Metadata +	Open data Linked data +	Big data Real-time (stream) data +	Intelligent (smart) data
Processes and activities, data lifecycle phases	Data generation Data transfer +	Innovation Data quality Data processing Data visualization +	Support Data sharing Participation Collaboration +	Transparency Openness Engagement (Re)use +	Data mashing Impact Decision-making Sustainability +	Resilience Interoperability Storytelling

Fig. 3. Conceptual model of the evolutionary generations of public data ecosystems resulting from the SLR.

Citizens and academia were involved in this generation of public data ecosystems to support innovation, improve data quality, or visualise the data to make them more accessible to users;

3. **the third generation** — "**public data sharing generation**" (typically from the 2000s to 2007/2009) is partly affected by the development of e-government and the spread of ICT and access to the Internet among citizens and businesses. To standardise public data and information, various initiatives, guidelines, and recommendations were developed, and some public organisations were obliged to publish public sector data online, including metadata. Data sharing, participation, and collaboration (and co-creation) are supported by governments;
4. **the fourth generation** — "**open (government) data generation**" (typically from 2007/2009 to 2013/2015) – this is the generation of the first open (government) data ecosystems, in which dynamic processes and activities oriented towards an increased transparency, openness, engagement, and (re)use were performed to develop new applications, services, and gain value from public data. Users and developers were key for this generation of ecosystems, and they were encouraged to improve their data-related competencies and skills;
5. **the fifth generation** — "**public data-driven generation**" (typically from 2013/2015 to 2020s) is the second generation of open (government) data ecosystems, which are characterised by new components, such as big and LOD and cloud computing. They supported reuse and linking of big data to improve decision-making, sustainability, and overall impact of public data in the society. This generation is also more open than the previous ones and involves new stakeholders such as international organisations, which results in external pressures on the development of the ecosystem;
6. **the sixth generation** — "**intelligent public data generation**" (typically from the 2020s to today) – this generation is represented by intelligent algorithms, machine learning, natural language processing tools and AI in a broader sense that change the work with public data, as well as how users consume and interact with them. Intelligent (smart) public data form this ecosystem to support its resilience and interoperability. Storytelling is the key way to bring large volumes and a variety of data closer to citizens and other stakeholders. New actors and roles, such as ecosystem managers, data curators, data consultants, also emerge to manage public data ecosystems more efficiently. However, this generation can be seen as a forward-looking generation, and while its characteristics are defined based on the SLR results, some of those studies were of rather conceptual nature setting an agenda for the public data ecosystems of a new generation, i.e. characteristics they are expected to meet.

During internal revision of the model and discussions among the eight authors, which included analysis of the model in terms of its compliance with the real-world, i.e., practical experience of the authors with public data ecosystems, e.g., of their countries, it has been found that while these generations are generally valid, some deviations from practical experience can be observed. First, some generations are valid, except for the periods mentioned above, which were identified from the literature and are not necessarily generalizable. Secondly, in some generations the characteristics mentioned for subsequent generations may also be present, but they are not included due to their lower level of impact on the overall public data ecosystem, i.e., they do not yet shape an ecosystem, but are rather characteristics that exist but without a significant effect. This was the case with a stakeholder "data user", which did not exist in the initial version of the model for the first generation of ecosystems, since the very first generation of public data ecosystems was focused on publishing data with limited focus on the data user and reuse by external parties. However, in this case, considering the discussion among the authors, it was decided to include this stakeholder as an "internal data user", since this stakeholder was part of the ecosystem, while the "external data user" — the user in a more traditional understanding, became a formative characteristic only for the second generation of public data ecosystems.

Thirdly, in some countries it may be that a particular generation of the public data ecosystem had most characteristics it should be characterised by, but one or several were missing, becoming of greater interest in subsequent years, when other characteristics would allow the public data ecosystem to be classified as the ecosystem of the next generation. For some countries, one or more characteristics are still missing, and consideration of their exclusion from the model was rejected by other authors because it was found to be relevant for all other cases with which the authors are familiar with, and in the end, the model conceptualise three decades of the research, i.e., the literature on this topic.

Fourthly, some generations – the first and the last – turned out to be the most difficult to assess their compliance with the real situation in practice. For the first generation of public data ecosystems, this was mainly due to limited or no transparency and awareness of the existence of public data at that point, as well as a very low level of its maturity, high decentralisation, and level of fragmentation of various systems that hardly constituted an ecosystem. For the latest, namely the sixth generation, this is due to the fact that it is based on the results of the SLR, where some of those resources can be characterised as calls for actions, and forward-looking ideas about what public data ecosystems are expected to look like in the future with a dominant focus on their resilience, sustainability and compliance with both user-centric design and emerging technologies. This is typically the case for theoretical and conceptual literature on the sustainable open data ecosystems, data spaces, and fair data ecosystems built around the concept of data republic. Therefore, we rather see this generation of public data ecosystems as a more provisionary and forward-looking generation of public data ecosystems.

5. Discussion and limitations

In this section, we first discuss how public data ecosystems have evolved in the literature, their value-adding elements, and generational shifts and identify future research directions as implied from our study. We then address the limitations of this study, some of which may also inform future research directions. Finally, we present practical and theoretical implications of this research.

5.1. Discussion

Data in all volumes, formats, from various stakeholders are forming public data ecosystems. However, not all data within these ecosystems get the attention they need in the context of public data lifecycle, which in turn affects the value creation from these data and their contributions to decision-making support. The proliferation of the Internet, cloud computing, and data analytics have significantly influenced the evolution of public data ecosystems, especially how the collection, storage, analysis, and sharing of larger and more complex datasets unfolds. Further, the emergence of the open data movement has changed how governments and organisations have embraced open data policies to make public information more accessible and transparent. The development of data governance frameworks and policies that guide the responsible and ethical use of public data shaped data governance practices. Citizens and communities have become active contributors to data collection and analysis, influencing decision-making processes.

An important requirement to achieve value-adding public data ecosystems is to know *what data public sector organisations and institutions have* and *how they should work with them to get value from them*. Considering the ecosystems perspective, which provides an approach to describe relationships between formative components and relationships, stakeholders, actors and their roles, data types, processes and activities and/or data lifecycle phases – it becomes imperative to discuss *how such a perspective influences the evolution over time* and *how public data ecosystems should be prepared for these changes*. According to Aaen et al. (2022), there will always be bright spots and dark spots in the value of the data ecosystem as it grows. In this regard, policymakers and managers need to carefully focus on aligning stakeholders, capabilities, and data, not only noting the benefits but also considering the risks of an evolving data ecosystem.

Social, psychological, and ethical aspects resulting from the dynamics of data ecosystems can also affect the development of the ecosystem. Public data ecosystems are also influenced by local culture, political institutions, and historical influences that form the institutional conditions in which ecosystems are located or in which they must be embedded. Such conditions shape the functioning of actors within the ecosystem and the arrangement of various ecosystem elements (Haak et al., 2018). This understanding is now more critical than ever, considering the current transformation of Society 4.0 into Society 5.0 (Fukuyama, 2018), which is expected to increase transparency and active participation in solving social challenges. Ensuring equal opportunities for all and integrating innovative technologies and society through public data ecosystems can be seen as prerequisites for this transition (Dewi et al., 2021; Nikiforova et al., 2023a).

International comparisons of public data ecosystems aid in discerning similarities and differences in countries' experiences, thereby contributing to a deeper understanding of the impact of data policies (Bates, 2014). Comparing and contrasting these ecosystems should facilitate identification of the key processes and resources, essential for their development. Furthermore, it is also valuable to explore the implications of different types of implementation experiences (Styrin et al., 2017). An established and mature e-government infrastructure further enhances the quality of this ecosystem (Ahn and Chu, 2021). Another aspect that is important for the development of the public data ecosystems is the extent of interaction between involved stakeholders, the clarity of processes occurring in the ecosystem, and their contribution to value creation. According to Jetzek (2017), all actors, especially data providers and data users, must have a reasonable understanding of prospective future values and what is required to maximise those benefits. Equally crucial is to consider the incentives for various stakeholders to share (or not share) the data to promote the creation and evolution of a data ecosystem (Van den Homberg and Susha, 2018).

Although data ecosystems are considered as socio-technical systems (Gelhaar and Otto, 2020; Oliveira et al., 2019), the technical perspective and its associated requirements are often overlooked in the evolution (lifecycle) of these ecosystems. Specifically, attributes such as scalability, extensibility and other quality indicators are not sufficiently represented. Additionally, decisions and constraints regarding technologies, protocols, standards, and contractors, and how they are implemented and evolved, can lead to infrastructure fragmentation (Kitchin and Moore-Cherry, 2021). In addition, there are many challenges regarding infrastructure, governance, systems engineering, and human-centricity, i.e., trusted data platforms, ecosystem data governance, and incrementally evolving systems engineering (Curry and Sheth, 2018).

We also found that public data ecosystems are built at different levels – country or national, city or local or municipal, international, or supranational. Each of the levels possesses its own distinct elements, boundaries, and environments, fostering the growth of the ecosystem, where more collaboration leads to higher operational efficiency, suggesting a positive scale effect (Clement et al., 2022; Janssen and Estevez, 2013). However, the current research suggests that as the local governments have limited resources to manage the ecosystem, increased collaboration is likely to be observed only up to a certain point (Botequilha-Leitão and Díaz-Varela, 2020). Many governments and public agencies still refuse or fail to share their data and thus are unable to utilise the offerings of data ecosystems. However, the actual sharing of data can also generate costs in terms of effort and time.

While previous research has made contributions to public data ecosystems, it often overlooks to theoretically integrate the various elements of these ecosystems and describe their evolution over different generations. Hence, this study extends the findings of earlier comprehensive review approaches towards different types of public data ecosystems such as geodata / spatial / geospatial data ecosystems (e.g., Coetzee and Wolff-Piggott, 2015; Oliveira and Lisboa Filho, 2015; Rajabifard et al., 2002), open (government) data ecosystems (e.g., Harrison et al., 2012; Heimstädt et al., 2014a), big data ecosystems (e.g., Munshi, 2018; Parida and Tripathi, 2018, Shah et al., 2020), Smart City and IoT data ecosystems (e.g., Curry and Ojo, 2020; Gupta et al. 2020; Kalaitzakis et al., 2019; Tron, 2020), domain-specific data ecosystems (e.g., Aaen et al., 2022; Gulson and Sellar, 2019; Haak et al., 2018; Hardy and Liu, 2022; Hoeyer, 2020; Park and Gil-Garcia, 2017; Ramalli and Pernici, 2022), stakeholder-centred data ecosystems / collaborative data partnerships (e.g., Otto and Jarke, 2019; Ruijer et al., 2023; Susha et al., 2023), policy and governance data ecosystems (Calzada and Almirall, 2020; Curry and Ojo, 2020; Hardy and Liu, 2022; Hoeyer, 2020; Hokka, 2022; Linåker and Runeson, 2021; Schneider and Comandé, 2022), and public data ecosystems in general (Gelhaar et al., 2021; Heinz et al., 2022; Lis and Otto, 2021; Oliveira et al., 2019). In addition, two rather emerging types of public data ecosystems were identified, namely fair data ecosystem (Calzati and van

Loenen, 2023) and public data space (Beverungen et al., 2022).

Since these studies lack a coherent linkage and the display of the relationships between different value-adding elements and how they contribute to forming of evolutionary generations, we offer a more holistic view on this topic. Given the increased momentum, the expanding breadth and velocity in the accumulation of public data ecosystems knowledge, this paper conducts a domain-based systematic review (Paul et al., 2021; Paul et al., 2023) synthesises existing, often disparate findings, explore trends and multivariate relationships between constructs by providing a comprehensive and unbiased summary of existing state-of-the-art findings, thereby drawing the big picture for a comprehensive and deeper understanding of the domain by both the academic, policy, and practitioner communities (Palmatier et al., 2018). It conceptually intersects with relevant information systems, systems theory, and public sector research, highlighting research gaps and explaining directions for future research (Moher et al., 2009; Paul and Criado, 2020; Paul et al., 2023; Snyder, 2019).

In this study, studies found within SLR are classified by topics they focused on and types of public data ecosystems they consider. These findings, obtained by summarising, organising, and conceptualising current research on public data ecosystems, have contributed to the development of understanding of the evolution of these ecosystems. As a result, we proposed the conceptual model that outlines the evolution of these ecosystems over time, from early stages focused on basic data infrastructure to more advanced stages leveraging emerging technologies, including key milestones that facilitate the transition from one generation of the ecosystem to the next. Each generation builds upon the previous one, incorporating new technologies, policies, and stakeholders, where the sixth generation is described as forward-looking and somewhat speculative, as it anticipates the future development of public data ecosystems based on current trends and conceptual studies. Consequently, we have identified future research directions. Although we briefly mentioned several potential directions earlier, the next section presents a structured and more detailed research agenda as implied from our research.

5.2. Future research agenda

Our findings from the SLR revealed limited research on (1) the diversity of different types of public data ecosystems and their impact on different groups of stakeholders; (2) governance models and frameworks, and their impact on the various processes and data requirements occurring in public data ecosystems; and (3) emerging technologies and how they shape the structure and processes of public data ecosystems.

Thus, we consider these topics to be particularly important for future research. We synthesised insights from the current literature on public data ecosystems and results obtained from answering the RQs, i.e., the typology of public data ecosystems (Fig. 1), the conceptual model of the public data ecosystem and its *meta*-characteristics (Fig. 2), and, in particular, the conceptual model of the evolutionary generations of public data ecosystems and its sixth forward-looking generation (Fig. 3) to identify the key future research avenues and corresponding subtopics. These future avenues can be classified into three main research streams: (a) the impact of public ecosystems on public value models; (b) technical and governance aspects in public data ecosystems; and (c) the impact of technological advances on the structure and processes of public data ecosystems.

5.2.1. Impact of public data ecosystems on public value models

Recent research shows the growing interest of cities, especially smart cities, in implementing effective and efficient **local government data ecosystems or smart city data ecosystems** (Bonina and Eaton, 2020; Liva et al., 2023; Lnenicka et al., 2024a, Ruijer et al., 2024; Wilson and Cong, 2021) to better understand the needs of citizens at lower levels of public administration. Better use of the vast amounts of data generated by cities, which often remain underutilised (Liva et al., 2023), can help local governments implement more open and collaborative governance models, as well as to provide citizen-centric services, improving both the urban planning and the efficient resource allocation (Lnenicka et al., 2022). Thus, future research could explore whether public data ecosystems have an impact on stakeholder participation in government and public decision-making and whether smart cities that implement smart city data ecosystems reach higher levels of citizen-centric services and higher levels of public value creation.

Nonetheless, the implementation of smart city data ecosystems at an urban scale can be difficult both due to the existing resources (technical, financial etc.) and due to the complex, holistic, and large volume of data and information that needs to be structured, processed, and distributed into these ecosystems. In this way, public administrations could find it more feasible to focus their efforts on implementing **domain-specific data ecosystems** in accordance with their strategic focus at the national, regional or city level, or even at sectoral or industry level (e.g., transportation or healthcare). The implementation of domain-specific public data ecosystems can be an objective of public administrations to tackle environmental, social, economic challenges to achieve the Sustainable Development Goals (SDGs) (Hein et al., 2023) or even as a complement to other global smart city data ecosystems to focus cities' attention on relevant urban challenges that need to be addressed. These domain-specific data ecosystems may include data as one of their core elements, but be complemented by other equally important elements such as processes (sustainability, innovation, and interoperability) and stakeholders (e.g., networks and partnerships), with the ultimate goal of creating public value. As such, future research directions could analyse empirical evidence of domain-specific data ecosystems regarding their potential ability to both meet the SDGs and solve urban challenges. Future research can explore *the different elements of these domain-specific data ecosystems and their impact on the process of creating public value*.

Effective implementation of public data ecosystems, in turn, requires **citizens' educational initiatives and training programs aimed at improving data literacy and skills** to develop a workforce capable of harnessing the potential of data ecosystems (Ansari et al., 2022; Gascó-Hernández et al., 2018; Lnenicka et al., 2024b; Santos-Hermosa et al., 2023)). As noted by Hokka (2022), in a data society, higher level of citizens' digital and data infrastructure literacy can help them clearly understand data and information from various digital sources, better exercise their rights, and to better support their active participation in society (Reggi et al., 2016; Visser, 2013), which may influence the implementation of participatory and collaborative governance models. In addition, data literacy plays an important role, not only in the use of data by either users or producers (Santos-Hermosa et al., 2023), but also in data governance (Clark and Albris, 2020; Sharp et al., 2022), opening the possibilities of embedding the potential of both OGD and open research data (Santos-Hermosa et al., 2023). Therefore, future research could examine *to what extent data literacy is present in citizens and other stakeholders in different contexts (smart cities, administrative cultures, specific domains), and at different administrative levels (national, regional, or local), as well as the influence of data literacy on data usage, on citizen engagement in public decisions and on data governance models*.

In addition, it is necessary to provide **training and support to frontline civil servants** to achieve sufficient levels of data literacy or even open data literacy (Lnenicka et al., 2024b; Loría-Solano and Raffaghelli, 2021) and knowledge in advanced analytics, including data science (Lnenicka and Komarkova, 2019; Umbach, 2022). These competencies are essential for data-driven decision-making in governments of all levels, including cities, which means understanding what data explains reality, knowing how to use data to inform practice, and understanding the broader implications of the datafication of society (Dingelstad et al., 2022). However, government efforts to train public administration staff are uneven. Thus, while some city governments are investing heavily in organisation-wide knowledge base and literacy, others barely put data/algorithms on the agenda (Fest et al., 2022). Moreover, such data literacy training becomes even more relevant in the new age of emerging technologies such as AI (European Union, 2022; Stankovich et al., 2023). In this regard, future research directions could examine *the relationship between investments in the digital literacy of public servants and the wider adoption of emerging technologies in both public services and decision-making processes*. In addition, new research could be conducted to analyse *this relationship with both the higher quality of public services provided and the impact of public decision-making processes on a higher level of citizens' quality of life*.

Last, but not least important, another unsolved issue is how to provide **data literacy to citizens and other stakeholders** (Salomão Filho et al., 2023). Although public administrations should be the responsible bodies for implementing data literacy programs, the way such programs are implemented can take many different forms. Some research streams point to educational institutions as being responsible for providing data literacy training to students in public education (Hilger et al., 2023) or even national government programs using various tools for providing data literacy training, for example, government projects using workshops to solve quizzes, riddles, and questions about privacy and personal data protection (Seymoens et al., 2020), digital innovation competitions, hackathons or datathons for civil society or specific target audience, such as enterprises and start-ups (Kitsios and Kamariotou, 2023) or schools (Nikiforova, 2022; Wolff et al., 2019). Others point to establishing connections with other anchor institutions, such as NGOs (Salomão Filho et al., 2023) or public libraries (Gasco-Hernandez et al., 2022), as complementary ways to teach data literacy. Even others point to the need for joint educational programs using public–private funding but with continued political support (Calzati and van Loenen, 2023). Finally, Solano et al. (2023) point out that open data activities also create opportunities to develop citizens' technical data literacy, allowing them to understand and interact with data-driven decision-making processes. However, little attention has so far been paid to the effectiveness of various methods for acquiring data literacy. Thus, future research could analyse *the methods used to acquire data literacy and their impact on both data use and data governance models*.

5.2.2. Technical and governance aspects in public data ecosystems

Another research stream focuses on analysing the technical and governance aspects of public data ecosystems. First, global challenges, such as those related to climate change or pandemic (and crises as a whole), have highlighted the need for **international cross-border cooperation and collaboration in data sharing** (Barthelemy et al., 2022). The increasing range of data coming from different sources increases the complexity of data ecosystems and requires a focus on data interoperability as a key element for effectively harnessing the value of these data (European Commission, 2020), especially regarding their reuse (European Parliament, 2019).

However, although **the standardisation for the data economy and data interoperability** has gained momentum in recent years and raises many issues to be addressed, until now public data interoperability has received relatively little attention, especially in the field of open data ecosystems research (Ali et al., 2022). Among other reasons, the convergence between technological development and interoperability has not occurred completely in parallel, mainly due to the many complex aspects, social, economic, and political dimensions involved (Serrano et al., 2022), which made interoperability aspects fall behind. In this regard, one of the key areas to be analysed is *how organisations can ensure that their internal IT landscape (data requirements and processes) is compatible with the data infrastructure*. Therefore, future research should focus on *how organisations take actions at* ***the four layers of the interoperability, namely technical, semantic, organisational, and legal layers*** (European Commission, 2017). Given the growing need for data exchange in accordance with various documents, such as European Commission (2017), the European strategy for data and the Data Governance Act (European Parliament, 2022), another future research trend can be an *analysis of the current impact of data interoperability issues on trust in data sharing and how this factor influences the use of data for decision-making processes.*

Finally, since public data ecosystems are built for a great variety of stakeholders with multiple and potentially conflicting interests, research can delve deeper into the analysis of digital transformation strategies and programs that must be developed to incorporate all these interests into the design of these ecosystems (Beverungen et al., 2022; Dawes et al., 2016). This means not only examining *how to fairly integrate all the competing and conflicting interests into the public data ecosystem*, but also, given the importance attributed to user interface design, *empirically evaluating the impact of design elements on user experience and engagement, data utilisation, and ultimately ecosystem sustainability and resilience* (Lee, 2024; Oliveira et al., 2019).

A second stream of research on the technical aspects of public data ecosystems that builds on our findings concerns **governance mechanisms and their social implications in data ecosystems**. To overcome adoption barriers in public data ecosystems, data governance elements, data audits, data access, or data licensing (Lee, 2014), are needed to distribute rights and responsibilities for data-related decisions across organisations in a multi-stakeholder scheme (Khatri and Brown, 2010). Governance mechanisms aimed at creating effective and democratic data ecosystems must be adapted to these multi-stakeholder policy schemes (Calzada and Almirall, 2020), resulting in governance arrangements that cross traditional policy scales and sectors, even though to the fact that governments and other stakeholders continue to be central to such developments (Hardy and Liu, 2022; Calzati and van Loenen, 2023). In this regard, since national data policies currently do not prioritise the creation of sector-specific data requirements or the establishment of sector-based ecosystems (Hrustek et al., 2023), future research should study *how these sector-specific requirements affect public data ecosystems*?

Governance mechanisms are context-dependent (Calzati and van Loenen, 2023; Hardy and Liu, 2022; Heinz et al., 2022), making it imperative that future research explores not only the different governance mechanisms implemented in public data ecosystems around the world, but also **the underlying factors that could shape these governance mechanisms and their effectiveness to enhance social implications** such as data reliability, the social and economic value of datasets, including their quality and the impact of high-value datasets (Geisler et al., 2022; Nikiforova et al., 2023b). On the other hand, public data ecosystems offer the opportunity to engage many different stakeholders, such as retailers (Beverungen et al., 2022) or data stewards (Wilkinson et al., 2022), opening up new opportunities regarding the roles played by these stakeholders in the public data ecosystem. These roles can impact the performance (Welle Donker and van Loenen, 2017), the structure (Susha et al., 2023) or even the data governance mechanisms of public data ecosystems. Therefore, future research directions should further explore *how public data ecosystem structures are designed according to the role of different stakeholders in the ecosystem and how this impacts the performance and governance mechanisms*. Since data governance mechanisms are composed of structural, procedural, and relational mechanisms (Abraham et al., 2019; Borgman et al., 2016), future research directions could analyse *how each of these mechanisms may be affected by the roles that are played by various stakeholders in this ecosystem*?

In addition, until now, previous research has focused on methods for the indirect assessment of data governance, ignoring the role of stakeholders, which may lead to a lower trust, motivation, and participation of these stakeholders (Lee et al., 2018; Reggi et al. 2016). **Data governance frameworks** are therefore needed to define direct assessment of data governance within public data ecosystems. In this regard, future research could *design these frameworks to allow these ecosystems to directly assess data governance requirements from stakeholders*.

Finally, the third research stream on the technical aspects of public data ecosystems concerns **the ethical implications of these ecosystems** caused by their structures and processes (Heinz et al., 2022; Rantanen et al., 2019). According to our findings, future research should *analyse these structures and processes to promote a better understanding and mitigation of the potential risks associated with these ecosystems while maximising their social benefits*. The goal of this research direction is to *build efficient ecosystems that strive to achieve higher social benefits while limiting the potential risks associated with the structure and processes within them*.

5.2.3. *Impact of technological advances on the public data ecosystems' structure and processes*

As previously noted, public data ecosystems are multidimensional (Beverungen et al., 2022; Hrustek et al., 2022; Ruijer et al., 2023) and dynamic in their nature (Oliveira and Lóscio, 2018; Wilson, 2019), evolving over time to respond to changes in technology, policy, and user needs (Jetzek, 2017). The development of **emerging technologies is disrupting the formation of public data ecosystems** and requires a **re-evaluation of effective data governance models** (Aaen et al., 2022; Tan, 2022). Big data analytics was just the beginning (Lnenicka and Komarkova, 2019; Zillner et al., 2021); currently, AI, including intelligent algorithms, natural language processing tools, machine learning, deep learning, and newer trends such as Generative AI and large language models, are increasingly becoming part of the public data ecosystems landscape. However, technological advances are not only a driving force for improving these ecosystems, but also a limiting factor that requires further exploration of technical aspects and system development as technology continues to rapidly evolve.

The adoption of AI in public data ecosystems could impact several areas that are likely to receive particular interest in future research, namely, improving data findability, data discoverability and data accessibility (Ahmed, 2023), moving towards data governance (Tan, 2022) to optimise the use of data. In the context of the sixth generation of public data ecosystems, AI and Generative AI are poised to assume pivotal roles in both automation and augmentation. This encompasses a spectrum of functionalities including processing tools, recommendation tools, and analysis tools (automation), as well as individualised, enriching, and ideation feedback mechanisms (augmentation). These advancements herald a comprehensive AI-assistance scope, spanning individual, peer-to-peer, and collective levels of interaction, aligning with the discourse outlined by Bono Rossello et al. (2024) on the "AI-enhanced online ideation", as well as some evidences of chatbots integrated into open data ecosystem (Cortés-Cediel et al., 2023; Porreca et al., 2018).

However, the use of AI in public data ecosystems also poses technical and data interoperability challenges, such as those related to the semantic interoperability, the vast quantities of data consumed by AI, or the cautions for curators and users of public datasets regarding the automated decision-making enabled by these technologies (Robinson and Scassa, 2022). Thus, in the context of emerging technologies in public data ecosystems, two general directions for future research streams are (1) the **opportunities** presented by these advances, covering both *their direct integration into ecosystems* and *their peripheral use by stakeholders and actors*, and (2) the **risks** associated with them, which must be carefully considered in this new era, which includes, among others, the need to *consider the ethical and legal considerations that accompany the integration of these technologies*.

The examination and implementation of these technological advances are expected to be closely intertwined with the study of their **adoption or resistance**. To conduct this analysis, future research is expected to examine *drivers and inhibitors for AI adoption in public data ecosystems*, highlighting not only benefits and challenges of AI adoption, but also the contextual, institutional, and organisational factors that should be considered when deciding to adopt it. This will likely lead to **the development of new theories and models**

regarding technology acceptance or resistance to adoption within these ecosystems by owners/managers and users of these data ecosystems.

In addition, the integration of AI and Generative AI into the data ecosystem leads to several critical considerations that need to be considered regarding data due, i.e., ensuring that the data being made open is **AI-ready**, also in line with (Open Data Institute, 2023). Since the AI and Generative AI models rely on the data they are trained on, **data quality** is paramount, encompassing not only completeness, but also other data quality dimensions, including accuracy, reliability, and comprehensive coverage of relevant information, including metadata and their quality, as well as associated processes of data preparation and pre-processing, involving cleaning, normalisation, and standardisation to ensure compatibility with (generative) AI technologies. Paradoxically, while AI and Generative AI may contribute to enhancing data quality and metadata, they inherently rely on their pre-existence. Thus, data must meet these criteria independently to optimise their utilisation. Thus, future research may focus on **the impact of AI and Generative AI on data requirements and quality**. While some of the above are noteworthy for the previous data ecosystem generations, the level of their importance is expected to increase even more within the sixth generation identified in our research.

On the other hand, in recent years, the focus on the predictive capabilities of AI algorithms has often overshadowed the importance of understanding the underlying mechanisms and AI decision interpretations (Dwivedi et al, 2023). But, within public data ecosystems, ensuring **transparency, trustworthiness, fairness, interpretability, and explainability of AI systems** (known as explainable AI (XAI)) become essential for fostering understandable, accountable, and reliable AI-driven decision-making processes, thereby enhancing trust among users and stakeholders (Niermann and Kügler, 2021). Overall, the role of XAI in enhancing the transparency, interpretability, and trustworthiness of AI decision-making in these ecosystems presents a promising avenue for future research, potentially addressing ethical and legal concerns. To achieve this goal, it would be valuable to examine *the stakeholders' perceptions of the impact of the ethical and legal implications of AI in public data ecosystems*, offering a comprehensive understanding of this topic.

5.3. Limitations

Our study has certain limitations. When combining data from research databases, there is a possibility that time and information technology constraints will prevent the full range of relevant research from being represented. Also, even though this research relied on a large sample of papers from respected and comprehensive databases (Scopus, WoS, and DGRL), relevant research may exist outside these queried sources. Moreover, our keyword-based search approach may have inadvertently omitted relevant studies, due to variations in authors' keyword selections. In addition, because our final sample consisted only of studies in English, we may have overlooked some potentially important research in other languages. Some studies also had restricted access to full texts, which may have led to the exclusion of some important studies. SLRs are also based on existing literature up to a certain cutoff date, and new research continues to be published after that date. This can cause reviews to become outdated after publication, potentially leading to recommendations that do not reflect the most recent evidence. Although we conducted a literature search twice in this study to ensure that the corpus is as complete and up-to-date as possible, the above limitation remains in effect with respect to the date the article becomes publicly available. Additionally, even with strict protocols in place, as is the case for this study, reviewer bias can still influence the selection and interpretation of studies. Personal beliefs, preferences, or preconceptions may unconsciously influence the review process.

Last but not the least, the limitation lies in the methodology itself, since although the SLR is a widely accepted, adopted and relevant methodology for this type of study, it can be seen as limited as it only covers knowledge that has been published with potential oversight or delay in reflecting the actual real-world situation. Therefore, the proposed models are expected to be validated and refined through their examination in the context of real public data ecosystems, with the first steps in this direction - validating EMPDE already done in (Nikiforova et al., 2024).

5.4. Study implications

5.4.1. Theoretical implications

The findings of this study contribute to the current theoretical body of knowledge by summarising, structuring, and conceptualising 30 years of research in public data ecosystems, using an ecosystem perspective and an ecosystem approach to describe the key value-adding elements and formative characteristics that make up the public data ecosystems.

First, this study presents a typology of public data ecosystems and a conceptual model of elements and formative characteristics that contribute to value-adding public data ecosystems to guide researchers toward a better understanding of the phenomenon and to prevent replicative research that does not substantially advance knowledge, highlighting the research gaps and clarifying directions for future research to substantially advance knowledge. As such, this paper defines a core terminology for public data ecosystems, aiming to ensure consistent use by the research community, and to clarify between similar yet distinct concepts. This also includes establishing an understanding of the different types of public data ecosystems found in the literature and the nascent yet promising public data spaces and fair data ecosystem, which broaden the scope of the public data ecosystem. Each type of public data ecosystem is accompanied by a list of studies that have covered it, serving as a reference point for researchers in the area.

Second, this paper reflects the evolution of public data ecosystems, and their elements and formative characteristics, and the relationships between them, to provide guidance for ecosystem development and growth, including the milestones that triggered generational transitions. As a result, this study develops a conceptual model of the evolutionary generation of public data ecosystems, represented by six generations. We also found that evolution tended to be more static in the past, and conversely, more dynamic within contemporary ecosystems that have focused on modern technologies, engagement of stakeholders, and considering processes to support value (co–)creation and decision-making. This study integrates conceptual knowledge about the elements of public data ecosystems and their evolution, contributing to the understanding of the development and dynamics of these ecosystems.

Finally, considering the last of the identified generations – the sixth – is the forward-looking generation, this study lays the foundation for future research to identify the most prominent areas for joint contribution to the research and development of the next generation of public data ecosystems.

5.4.2. Practical implications

To improve and adapt public data ecosystems to changes arising from the development of the digital society and technologies, practitioners should reconsider the relationships between individual key value-adding elements, considering their strengths and influence on each other. In turn, the revision of current public data ecosystems requires an awareness of these, based on previous experience gained over the years.

In this study, we identified the most common elements and relationships between them within public data ecosystems, focusing on elements that must be inherent in the ecosystems and implemented at the proper level. Practice shows that public data ecosystems do not always comply with these elements, which may be due to a limited understanding of the above. Our findings are therefore relevant for practitioners developing and designing public data ecosystems, guiding practitioners in identifying and prioritising elements and fostering a comprehensive understanding of the intricate relationship between them.

These insights are relevant not only for designing ecosystems from scratch, but also for revising an existing system, i.e., setting an improvement agenda with an emphasis on the ecosystem as a whole, its components, the relationships between these components, and the processes occurring in the ecosystem. Our study facilitates the identification of value-adding elements, their interconnections, and the determination of generation of the ecosystem under, i.e., whether it aligns with the current paradigm of an up-to-date ecosystem, or requires revisions to increase its relevance to the current generation. Understanding these aspects enables the identification of missing or inadequate elements and informs evidence-based design interventions.

Finally, this study suggests directions for future research to further improve the practical understanding and application of public data ecosystems, advance knowledge, address real-world challenges and contribute to the practical application of scientific findings, especially within the forward-looking generation of public data ecosystems.

6. Conclusion

The aim of this study was to explore public data ecosystems and their evolution over the years (RQ1), identifying the key value-adding elements that constitute them, the mutual relationships between them, and their formative characteristics (RQ2), identifying the evolutionary generations and elements that could influence the transition from one stage of public data ecosystem generation to the next (RQ3). To achieve this aim and answer the three research questions, a SLR over the past 30 years was conducted, resulting in 148 studies that have been thoroughly examined, illuminating the current landscape of data ecosystems research, and pointing to directions for future research and application. As a result of the SLR, we were able to design a typology of public data ecosystems consisting of thirteen types of public data ecosystems, all of which contribute to expanding the boundaries of public data ecosystems. Each type of public data ecosystem has been defined, emphasising its core elements, and defining relationships between these types, thereby answering RQ1.

We then examined and conceptualised the elements that make up public data ecosystems, consisting of three key groups of elements or *meta*-characteristics, namely, (1) data – data characteristics, data infrastructure, data lifecycle, data policy, (2) stakeholders – types of stakeholders, actors and roles, skills and competences, and (3) processes, which are then broken down into smaller elements, thereby answering RQ2.

Finally, we proposed a conceptual model of the evolutionary generations of public data ecosystems. Six generations of public data ecosystems were identified. The earlier generations are more static and data publishing-oriented, whereas later generations are more dynamic and value creation-oriented in later generations, with increasing emphasis on modern technologies, innovation creation (from the second generation), stakeholder engagement (from the third generation), and consideration of processes to support value (co–)creation and decision-making. Each generation was described based on key value-adding elements and formative characteristics, namely (1) components and relationships, (2) stakeholders, (3) actors and their roles, (4) data types, (5) processes, activities, and lifecycle phases, thereby answering RQ3.

To synthesise the results from this study, three avenues for future research were defined that can be classified into three main research streams: (a) the impact of public ecosystems on public value models; (b) technical and governance aspects in public data ecosystems; and (c) the impact of technological advances on the structure and processes of public data ecosystems.

This study intends to help practitioners in developing and designing public data ecosystems, determining the generation to which the current ecosystem belongs, and extracting the characteristics – value-adding elements and relationships between them – that the ecosystem is expected to be characterised by learning from defined generations. The findings also contribute to the current theoretical body of knowledge by structuring and conceptualising current research in the area, reflecting the evolution of ecosystems, and thereby moving towards defining public data ecosystems of the new generation.

CRediT authorship contribution statement

Martin Lnenicka: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anastasija Nikiforova:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mariusz Luterek:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Petar Milic:** Writing – original draft, Resources, Investigation, Conceptualization. **Daniel Rudmark:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Sebastian Neumaier:** Writing – original draft, Investigation, Conceptualization. **Karlo Kević:** Investigation. **Anneke Zuiderwijk:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Manuel Pedro Rodríguez Bolívar:** Writing – review & editing, Writing – original draft, Validation, Supervision, Formal analysis, Conceptualization.

Declaration of Generative AI use

The authors hereby discloses that ChtGPT-3.5 was used to improve the conciseness and clarity of selected sentences in this study. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data accompanying this article are available on Zenodo - https://zenodo.org/doi/10.5281/zenodo.13842001.

Acknowledgements

This research was funded by the Regional Government of Andalusia, Spain (Research project number C-SEJ-325-UGR23). The authors would like to thank the Ministry of Science, Technological Development and Innovation of the Republic of Serbia for funding the scientific research work, contract no. 451-03-65/2024-03/200155, realized by the Faculty of Technical Sciences in Kosovska Mitrovica, University of Pristina. This research is part of the Twinning Open Data Operational project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 85759.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tele.2024.102190

References

Aaen, J., Nielsen, J.A., Carugati, A., 2022. The dark side of data ecosystems: a longitudinal study of the DAMD project. Eur. J. Inf. Syst. 31 (3), 288–312.
Abraham, R., Schneider, J., Vom Brocke, J., 2019. Data governance: a conceptual framework, structured review, and research agenda. Int. J. Inf. Manag. 49, 424–438.
Ahmed, U. (2023). Reimagining open data ecosystems: a practical approach using AI, CI, and Knowledge Graphs. In B*IR 2023 Workshops and Doctoral Consortium, 22nd International Conference on Perspectives in Business Informatics Research (BIR 2023)* (pp. 235-249). CEUR-WS.org.
Ahn, M., Chu, S., 2021. In: What Matters in Maintaining Effective Open Government Data Systems? the Role of Government Managerial Capacity, and Political and Legal Environment. ACM, pp. 444–457.
Ali, M., Alexopoulos, C., Charalabidis, Y., 2022. Data interoperability and the open data ecosystem: roles and research areas. In: In *Book of Abstracts – ICOD 2022*, pp. 74–80.
Ansari, B., Barati, M., Martin, E.G., 2022. Enhancing the usability and usefulness of open government data: a comprehensive review of the state of open government data visualization research. Gov. Inf. Q. 39 (1), 101657.
Azkan, C., Möller, F., Meisel, L., Otto, B., 2020. Service dominant logic perspective on data ecosystems - a case study based morphology. In: In *Proceedings of the 28th European Conference on Information Systems (ECIS)*, p. (p. 65)..
Bailey, K.D., 1994. Typologies and Taxonomies: an Introduction to Classification Techniques, Vol. 102. Sage.
Barthelemy, F., Cochez, M., Dimitriadis, I., Karim, N., Loutas, N., Magnisalis, I., Comet, L.M., Peristeras, V., Wyns, B., 2022. Towards a standard-based open data ecosystem: analysis of DCAT-AP use at national and European level. Electronic Government, an International Journal 18 (2), 137–180.
Bates, J., 2014. The strategic importance of information policy for the contemporary neoliberal state: the case of Open Government Data in the United Kingdom. Gov. Inf. Q. 31 (3), 388–395.
Beverungen, D., Hess, T., Köster, A., Lehrer, C., 2022. From private digital platforms to public data spaces: implications for the digital transformation. Electron. Mark. 32 (2), 493–501.
Bonina, C., Eaton, B., 2020. Cultivating open government data platform ecosystems through governance: lessons from Buenos Aires, Mexico City and Montevideo. Gov. Inf. Q. 37 (3), 101479.
Bono Rossello, N., Simonofski, A., Clarinval, A., Castiaux, A., 2024. In: A Typology for AI-Enhanced Online Ideation: Application to Digital Participation Platforms. ScholarSpace, pp. 1850–1859.
Borgman, H., Heier, H., Bahli, B., Boekamp, T., 2016. In: Dotting the i and Crossing (out) the t in It Governance: New Challenges for Information Governance. IEEE, pp. 4901–4909.
Botequilha-Leitão, A., Díaz-Varela, E.R., 2020. Performance based planning of complex urban social-ecological systems: the quest for sustainability through the promotion of resilience. Sustain. Cities Soc. 56, 102089.
Bowker, G.C., 2000. Biodiversity datadiversity. Soc. Stud. Sci. 30 (5), 643–683.

Calzada, I., Almirall, E., 2020. Data ecosystems for protecting European citizens' digital rights. Transforming Government: People, Process and Policy 14 (2), 133–147.
Calzati, S., van Loenen, B., 2023. A fourth way to the digital transformation: the data republic as a fair data ecosystem. Data & Policy 5, e21.
Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E., 2018. Open data infrastructures. In: The World of Open Data. Springer, Cham, pp. 95–113.
Clark, N., Albris, K., 2020. In the interest (s) of many: Governing data in crises. Politics and Governance 8 (4), 421–431.
Clement, J., Manjon, M., Crutzen, N., 2022. Factors for collaboration amongst smart city stakeholders: a local government perspective. Gov. Inf. Q. 39 (4), 101746.
Coetzee, S., & Wolff-Piggott, B. (2015). A review of SDI literature: searching for signs of inverse infrastructures. In *Cartography - Maps Connecting the World* (pp. 113-127). Springer, Cham.
Commission, E., 2020. Public sector information, open data and big data (RP2020). Available from: https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/public-sector-information-open-data-and-big-data-rp2020.
European Commission (2017). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. European Interoperability Framework – Implementation Strategy. COM(2017) 134.
Cortés-Cediel, M.E., Segura-Tinoco, A., Cantador, I., Bolívar, M.P.R., 2023. Trends and challenges of e-government chatbots: advances in exploring open government data and citizen participation content. Gov. Inf. Q. 40 (4), 101877.
Crompvoets, J., Vancauwenberghe, G., Bouckaert, G., Vandenbroucke, D., 2011. Practices to develop spatial data infrastructures: exploring the contribution to e-government. In: Practical Studies in E-Government. Springer, New York, NY, pp. 229–248.
Curry, E., Ojo, A., 2020. Enabling knowledge flows in an intelligent systems data ecosystem. In: Real-Time Linked Dataspaces. Springer, Cham, pp. 15–43.
Curry, E., Sheth, A., 2018. Next-generation smart environments: from system of systems to data ecosystems. IEEE Intell. Syst. 33 (3), 69–76.
Danneels, L., Viaene, S., Van den Bergh, J., 2017. Open data platforms: discussing alternative knowledge epistemologies. Gov. Inf. Q. 34 (3), 365–378.
Davies, T., 2011. Open data: infrastructures and ecosystems. Unpublished paper, University of Southampton.
Dawes, S.S., Vidiasova, L., Parkhimovich, O., 2016. Planning and designing open government data programs: an ecosystem approach. Gov. Inf. Q. 33 (1), 15–27.
Demchenko, Y., De Laat, C., Membrey, P., 2014. In: Defining Architecture Components of the Big Data Ecosystem. IEEE, pp. 104–112.
Dewi, V., Amelia, A., Mursitama, T.N., 2021. In: Blockchain Technologies and Factor of ERP Implementation in Making Infrastucture for Society 5.0. IEEE, pp. 1–5.
D'Hauwers, R., Walravens, N., Ballon, P., 2022. Data ecosystem business models: value and control in data ecosystems. Journal of Business Models 10 (2), 1–30.
Ding, L., Peristeras, V., Hausenblas, M., 2012. Linked open government data [Guest editors' introduction]. IEEE Intell. Syst. 27 (3), 11–15.
Dingelstad, J., Borst, R.T., Meijer, A., 2022. Hybrid data competencies for municipal civil servants: an empirical analysis of the required competencies for data-driven decision-making. Public Personnel Management 51 (4), 458–490.
Diran, D., Hoppe, T., Ubacht, J., Slob, A., Blok, K., 2020. A data ecosystem for data-driven thermal energy transition: reflection on current practice and suggestions for re-design. Energies 13 (2), 444.
Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., Ranjan, R., 2023. Explainable AI (XAI): core ideas, techniques, and solutions. ACM Comput. Surv. 55 (9), 1–33.
Elwood, S., 2008. Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. Int. J. Geogr. Inf. Sci. 22 (1), 71–90.
Estermann, B., Fraefel, M., Neuroni, A.C., Vogel, J., 2018. Conceptualizing a national data infrastructure for Switzerland. Information Polity 23 (1), 43–65.
Fest, I., Wieringa, M., Wagner, B., 2022. Paper vs. practice: how legal and ethical frameworks influence public sector data professionals in the Netherlands. Patterns 3 (10), 100604.
Fukuyama, M., 2018. Society 5.0: Aiming for a new human-centered society. Japan Spotlight 27 (5), 47–50.
Gama, K., & Lóscio, B. F. (2014). Towards ecosystems based on open data as a service. In *Proceedings of the 16th International Conference on Enterprise Information Systems (ICEIS-2014)* (pp. 659-664). Scitepress.
Gascó-Hernández, M., Martin, E.G., Reggi, L., Pyo, S., Luna-Reyes, L.F., 2018. Promoting the use of open government data: cases of training and engagement. G*Overnment Information Quarterly* 35 (2), 233–242.
Gasco-Hernandez, M., Yerden, X., Burke, G.B., Gil-Garcia, J.R., 2022. The potential role of public libraries in a quadruple helix model of "Smart City" development: lessons from Chattanooga. Tennessee. *the Library Quarterly* 92 (4), 419–436.
Geisler, S., Vidal, M.E., Cappiello, C., Lóscio, B.F., Gal, A., Jarke, M., Rehof, J., 2022. Knowledge-driven data ecosystems toward data transparency. ACM Journal of Data and Information Quality (JDIQ) 14 (1), 1–12.
Gelhaar, J., & Otto, B. (2020). Challenges in the emergence of data ecosystems. In *PACIS 2020 Proceedings* (p. 175). AIS.
Gelhaar, J., Groß, T., & Otto, B. (2021). A taxonomy for data ecosystems. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 6113-6122). ACM.
Grus, L., Crompvoets, J., Bregt, A.K., 2010. Spatial data infrastructures as complex adaptive systems. Int. J. Geogr. Inf. Sci. 24 (3), 439–463.
Gulson, K.N., Sellar, S., 2019. Emerging data infrastructures and the new topologies of education policy. Environment and Planning d: Society and Space 37 (2), 350–366.
Gupta, A., Panagiotopoulos, P., Bowen, F., 2020. An orchestration approach to smart city data ecosystems. Technol. Forecast. Soc. Chang. 153, 119929.
Haak, E., Ubacht, J., Van den Homberg, M., Cunningham, S., Van den Walle, B., 2018. In: A Framework for Strengthening Data Ecosystems to Serve Humanitarian Purposes. ACM, pp. 1–9.
Hardy, I., Liu, S., 2022. Complex connectivities: policy networks, data infrastructures and vocational education reform in China. Int. J. Educ. Res. 115, 102045.
Harrison, T.M., Pardo, T.A., Cook, M., 2012. Creating open government ecosystems: a research and development agenda. Future Internet 4 (4), 900–928.
Heimstädt, M., Saunderson, F., & Heath, T. (2014a). Conceptualizing open data ecosystems: a timeline analysis of open data development in the UK. In *CeDEM14: Conference for E-Democracy an Open Government* (pp. 245-255). MV-Verlag.
Heimstädt, M., Saunderson, F., Heath, T., 2014b. From toddler to teen: growth of an open data ecosystem. JeDEM-eJournal of eDemocracy and Open Government 6 (2), 123–135.
Hein, A., Engert, M., Ryu, S., Schaffer, N., Hermes, S., Krcmar, H., 2023. Building open government data platform ecosystems: a dynamic development approach that engages users from the start. Gov. Inf. Q. 40 (4), 101878.
Heinz, D., Benz, C., Fassnacht, M., & Satzger, G. (2022). Past, present and future of data ecosystems research: a systematic literature review. In *PACIS 2022 Proceedings* (p. 46). AIS.
Hilger, K., Gamböck-Strätz, J., Biedermann, C., 2023. In: Getting Data Literacy Education Ready for the Future. Springer Nature Switzerland, Cham, pp. 482–494.
Hoeyer, K., 2020. Data promiscuity: How the public–private distinction shaped digital data infrastructures and notions of privacy, 7. Humanities and Social Sciences Communications, pp. 1–8.
Hokka, J., 2022. In: Welfare Data Society? Critical Evaluation of the Possibilities of Developing Data Infrastructure Literacy from User Data Workshops to Public Service Media. Palgrave Macmillan, Cham, pp. 267–294.
Hrustek, L., Mekovec, R., & Alexopolus, C. (2023). Concept for an open data ecosystem to build a powerful data environment. In *Information Systems: 19th European, Mediterranean, and Middle Eastern Conference, EMCIS 2022* (pp. 251-263). Cham, Springer Nature Switzerland.
Hrustek, L., Tomičić Furjan, M., Varga, F., Džidić, A., van Loenen, B., Šalamon, D., 2022. Towards digital innovation: stakeholder interactions in agricultural data ecosystem in Croatia. Interdisciplinary Description of Complex Systems: INDECS 20 (2), 189–207.
Janowski, T., 2015. Digital government evolution: from transformation to contextualization. Gov. Inf. Q. 32 (3), 221–236.
Janssen, M., Estevez, E., 2013. Lean government and platform-based governance – doing more with less. Gov. Inf. Q. 30, S1–S8.
Jean-Quartier, C., Rey Mazón, M., Lovrić, M., Stryeck, S., 2022. Collaborative data use between private and public stakeholders – a regional case study. Data 7 (2), 20.
Jetzek, T., 2017. Innovation in the open data ecosystem: exploring the role of real options thinking and multi-sided platforms for sustainable value generation through open data. In: Analytics, Innovation, and Excellence-Driven Enterprise Sustainability. Palgrave Macmillan, New York, pp. 137–168.

Kalaitzakis, M., Bouloukakis, M., Charalampidis, P., Dimitrakis, M., Drossis, G., Fragkiadakis, A., ... & Stephanidis, C. (2019). Building a smart city ecosystem for third party innovation in the city of Heraklion. In *Mediterranean Cities and Island Communities* (pp. 19-56). Springer, Cham.

Kalliola, R., Käyhkö, N., Jokela, S., 2019. Adaptive development of a regional spatial data infrastructure facing local prospects and socio-technological trends. Bulletin of Geography. Socio-Economic Series 44, 73–80.

Khatri, V., Brown, C.V., 2010. Designing data governance. Commun. ACM 53 (1), 148–152.

Kitchenham, B., 2004. Procedures for performing systematic reviews. Keele University Technical Report 2004, 1–28.

Kitchin, R., Moore-Cherry, N., 2021. Fragmented governance, the urban data ecosystem and smart city-regions: the case of Metropolitan Boston. Reg. Stud. 55 (12), 1913–1923.

Kitsios, F., Kamariotou, M., 2023. Digital innovation and entrepreneurship transformation through open data hackathons: design strategies for successful start-up settings. Int. J. Inf. Manag. 69, 102472.

Klievink, B., Neuroni, A., Fraefel, M., & Zuiderwijk, A. (2017). Digital strategies in action: a comparative analysis of national data infrastructure development. In *Proceedings of the 18th Annual International Conference on Digital Government Research* (pp. 129-138). ACM.

Kok, B., Van Loenen, B., 2005. How to assess the success of national spatial data infrastructures? Comput. Environ. Urban Syst. 29 (6), 699–717.

Kontokosta, C.E., 2013. Energy disclosure, market behavior, and the building data ecosystem. Ann. N. Y. Acad. Sci. 1295 (1), 34–43.

Lee, P.C., 2024. Towards a more resilient ecosystem: Case study of open government data in Taiwan. Journal of Library and Information Studies 22 (1), 1–29.

Lee, S.U., Zhu, L., Jeffery, R., 2018. In: Designing Data Governance in Platform Ecosystems. IEEE, pp. 5014–5023.

Lee, D. (2014). Building an open data ecosystem: an Irish experience. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance* (pp. 351-360). ACM.

Lienert, C., Meier, S., 2014. Environmental data visualisation EnVIS–linking real-time sensor data to spatial data infrastructures for web-based visualisation. In: Thematic Cartography for the Society. Springer, Cham, pp. 293–304.

Linåker, J., Runeson, P., 2020. In: Collaboration in Open Government Data Ecosystems: Open Cross-Sector Sharing and Co-Development of Data and Software. Springer, Cham, pp. 290–303.

Linåker, J., Runeson, P., 2021. How to enable collaboration in open government data ecosystems: a public platform provider's perspective. JeDEM-eJournal of eDemocracy and Open Government 13 (1), 1–30.

Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 6067-6076). ACM.

Liva, G., Micheli, M., Schade, S., Kotsev, A., Gori, M., Codagnone, C., 2023. City data ecosystems between theory and practice: a qualitative exploratory study in seven European cities. Data & Policy 5, e17.

Lnenicka, M., Komarkova, J., 2019. Big and open linked data analytics ecosystem: theoretical background and essential elements. Gov. Inf. Q. 36 (1), 129–144.

Lnenicka, M., Nikiforova, A., Luterek, M., Azeroual, O., Ukpabi, D., Valtenbergs, V., Machova, R., 2022. Transparency of open data ecosystems in smart cities: definition and assessment of the maturity of transparency in 22 smart cities. Sustain. Cities Soc. 82, 103906.

Lnenicka, M., Nikiforova, A., Clarinval, A., Luterek, M., Rudmark, D., Neumaier, S., Kević, K., Rodríguez Bolívar, M.P., 2024a. Sustainable open data ecosystems in smart cities: a platform theory-based analysis of 19 European cities. Cities 148, 104851.

Lnenicka, M., Nikiforova, A., Luterek, M., Milic, P., Rudmark, D., Neumaier, S., Santoro, C., Casiano Flores, C., Janssen, M., Rodríguez Bolívar, M.P., 2024b. Identifying patterns and recommendations of and for sustainable open data initiatives: a benchmarking-driven analysis of open government data initiatives among European countries. Gov. Inf. Q. 41 (1), 101898.

Loría-Solano, E., & Raffaghelli, J. E. (2021). Competency model for open data literacy in professional learning within the context of Open Government Data (OGD). In *EC-TEL (Doctoral Consortium)* (pp. 102-112).

Martin, S., Turki, S., Renault, S., 2017. In: Open Data Ecosystems: Introducing the Stimulator Function. Springer International Publishing, Cham, pp. 49–63.

McBride, K., Olesk, M., Kütt, A., Shysh, D., 2020. Systemic change, open data ecosystem performance improvements, and empirical insights from Estonia: a country-level action research study. Information Polity 25 (3), 377–402.

McLaughlin, J., Nichols, S., 1994. Developing a national spatial data infrastructure. J. Surv. Eng. 120 (2), 62–76.

McLeod, M., McNaughton, M., 2016. Mapping an emergent open data ecosystem. J. Community Inform. 12 (2), 26–46.

McLoughlin, I., McNicoll, Y., Cornford, J., Davenport, S., 2019. Data-driven innovation in the social sector in Australasia – data ecosystems and interpretive communities. Public Money & Management 39 (5), 327–335.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151 (4), 264–269.

Mulder, A.E., Wiersma, G., Van Loenen, B., 2020. Status of national open spatial data infrastructures: a comparison across continents. International Journal of Spatial Data Infrastructures Research 15, 56–87.

Munshi, U.M., 2018. Data science landscape: tracking the ecosystem. In: Data Science Landscape. Springer, Singapore, pp. 1–31.

Nardi, B.A., O'Day, V.L., 1999. Information Ecologies: Using Technology with Heart. MA, MIT Press, Cambridge.

Nickerson, R.C., Varshney, U., Muntermann, J., 2013. A method for taxonomy development and its application in information systems. Eur. J. Inf. Syst. 22 (3), 336–359.

Niermann, C., Kügler, D., 2021. Next generation intelligent cockpit–single-pilot operations' potential to increase flight safety. ECAC News 74, 25–28.

Nikiforova, A. (2022). Open data hackathon as a tool for increased engagement of Generation Z: to hack or not to hack?. In *International Conference on Electronic Governance with Emerging Technologies* (pp. 161-175). Springer Nature Switzerland, Cham.

Nikiforova, A., Flores, M.A.A., Lytras, M.D., 2023a. The role of open data in transforming the society to Society 5.0: a resource or a tool for SDG-compliant Smart Living?. In: Smart Cities and Digital Transformation: Empowering Communities, Limitless Innovation, Sustainable Development and the next Generation. Emerald Publishing Limited, pp. 219–252.

Nikiforova, A., Lnenicka, M., Milić, P., Luterek, M., & Rodríguez Bolívar, M. P. (2024). From the evolution of public data ecosystems to the evolving horizons of the forward-looking intelligent public data ecosystem empowered by emerging technologies. In *International Conference on Electronic Government* (pp. 402-418). Cham: Springer Nature Switzerland.

Nikiforova, A., Rizun, N., Ciesielska, M., Alexopoulos, C., & Miletić, A. (2023b). Towards High-Value Datasets determination for data-driven development: a systematic literature review. In *International Conference on Electronic Government* (pp. 211-229). Cham, Springer Nature Switzerland.

Noorman, M. (2017). Institutions in the data ecosystem: Actors in the public knowledge domain and in private data companies. In *Open Data and the Knowledge Society* (pp. 85-103). Amsterdam University Press.

Oliveira, I. L., & Lisboa Filho, J. (2015). A spatial data infrastructure review: sorting the actors and policies from enterprise viewpoint. In *Proceedings of the 17th International Conference on Enterprise Information Systems (ICEIS-2015)* (pp. 287-294).

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a Data Ecosystem? In: *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age - dgo '18* (pp. 1-9). New York, ACM.

Oliveira, M.I.S., Oliveira, L.E.R., Batista, M.G.R., Lóscio, B.F., 2018. In: Towards a Meta-Model for Data Ecosystems. ACM, pp. 1–10.

Oliveira, M.I.S., Lima, G.D.F.B., Lóscio, B.F., 2019. Investigations into data ecosystems: a systematic mapping study. Knowl. Inf. Syst. 61 (2), 589–630.

Open Data Institute, 2023. *Data-centric AI* [online]. Available at: https://theodi.org/insights/projects/data-centric-ai/.

Orenga-Roglá, S., Chalmeta, R., 2019. Framework for implementing a big data ecosystem in organizations. Commun. ACM 62 (1), 58–65.

Otto, B., Jarke, M., 2019. Designing a multi-sided data platform: findings from the International Data Spaces case. Electron. Mark. 29 (4), 561–580.

Palmatier, R.W., Houston, M.B., Hulland, J., 2018. Review articles: purpose, process, and structure. J. Acad. Mark. Sci. 46, 1–5.

Parida, P.K., Tripathi, S., 2018. Odisha spatial data infrastructure (OSDI) – Its data model, meta data and sharing policy. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 4 (5), 21–28.

Park, S., Gil-Garcia, J.R., 2017. In: Understanding Transparency and Accountability in Open Government Ecosystems: the Case of Health Data Visualizations in a State Government. ACM, pp. 39–47.

European Parliament (2019). *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information*. Official Journal of the European Union, L172/56.

European Parliament (2022). *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance)*. Official Journal of the European Union, L152/1.

Paul, J., Criado, A.R., 2020. The art of writing literature review: What do we know and what do we need to know? Int. Bus. Rev. 29 (4), 101717.

Paul, J., Lim, W.M., O'Cass, A., Hao, A.W., Bresciani, S., 2021. Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). Int. J. Consum. Stud. 45 (4), O1–O16.

Paul, J., Khatri, P., Kaur Duggal, H., 2023. Frameworks for developing impactful systematic literature reviews and theory building: What, Why and How? J. Decis. Syst. 1–14.

Petersen, S.A., Pourzolfaghar, Z., Alloush, I., Ahlers, D., Krogstie, J., Helfert, M., 2019. In: Value-Added Services, Virtual Enterprises and Data Spaces Inspired Enterprise Architecture for Smart Cities. Springer, Cham, pp. 393–402.

Pinto, V. A., & Parreiras, F. S. (2020). Towards a taxonomy for big data technological ecosystem. In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 2* (pp. 294-305). SCITEPRESS.

Porreca, S., Leotta, F., Mecella, M., Vassos, S., Catarci, T., 2018. Accessing government open data through chatbots. In: Current Trends in Web Engineering. Springer International Publishing, Cham, pp. 156–165.

Rajabifard, A., Feeney, M.E.F., Williamson, I.P., 2002. Future directions for SDI development. Int. J. Appl. Earth Obs. Geoinf. 4 (1), 11–22.

Ramalli, E., Pernici, B., 2022. From a prototype to a data ecosystem for experimental data and predictive models. In: In *Proc. of the First InternationalWorkshop on Data Ecosystems (DEco'22)*, pp. 18–26.

Rantanen, M.M., Hyrynsalmi, S., Hyrynsalmi, S.M., 2019. In: Towards Ethical Data Ecosystems: a Literature Study. IEEE, pp. 1–9.

Reggi, L., Dawes, S.S., 2022. Creating open government data ecosystems: network relations among governments, user communities, NGOs and the media. Gov. Inf. Q. 39 (2), 101675.

Reggi, L., Dawes, S., (2016). Open government data ecosystems: Linking transparency for innovation with transparency for participation and accountability. In EGOV,, 2016. International Conference on Electronic Government. Springer, Cham, pp. 74–86.

Robinson, P., Scassa, T., 2022. The future of open data. University of Ottawa Press.

Ruijer, E., Dingelstad, J., Meijer, A., 2023. Studying complex systems through design interventions probing open government data ecosystems in the Netherlands. Public Manag. Rev. 25 (1), 129–149.

Ruijer, E., Dymanus, C., van Kesteren, E.J., Boeschoten, L., Meijer, A., 2024. Open data work for empowered deliberative democracy: findings from a living lab study. Gov. Inf. Q. 41 (1), 101902.

Salomão Filho, A., Wasson, B., Lameras, P., Maloszek, R., 2023. Data literacy for citizenry: a few policy recommendations from a literature review. Edutec. Revista Electrónica De Tecnología Educativa 86, 6–21.

Santos-Hermosa, G., Quarati, A., Loría-Soriano, E., Raffaghelli, J.E., 2023. Why does open data get underused? A focus on the role of (open) data literacy. In: Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead. Springer International Publishing, Cham, pp. 145–177.

Schneider, G., Comandé, G., 2022. It's time: leveraging the GDPR to shift the balance towards research-friendly EU data spaces. Common Market Law Review 59 (3), 739–776.

Serrano, M., Curry, E., Walsh, R., Purtill, G., Soldatos, J., Ferraris, M., Troiano, E., 2022. Data space best practices for data interoperability in FinTechs. In: Data Spaces: Design, Deployment and Future Directions. Springer International Publishing, Cham, pp. 249–264.

Seymoens, T., Van Audenhove, L., Van den Broeck, W., Mariën, I., 2020. Data literacy on the road: setting up a large-scale data literacy initiative in the DataBuzz project. Journal of Media Literacy Education 12 (3), 102–119.

Shah, S.I.H., Peristeras, V., Magnisalis, I., 2020. Government (big) data ecosystem: definition, classification of actors, and their roles. International Journal of Computer and Information Engineering 14 (4), 102–114.

Shah, S.I.H., Peristeras, V., Magnisalis, I., 2021a. DaLiF: a data lifecycle framework for data-driven governments. Journal of Big Data 8 (1), 1–44.

Shah, S.I.H., Peristeras, V., Magnisalis, I., 2021b. Government big data ecosystem: definitions, types of data, actors, and roles and the impact in public administrations. ACM J. Data Inf. Qual. 13 (2), 1–25.

Shakeri, M.A., Alimohammadi, A., Sadeghi-Niaraki, A.A., Alesheikh,, 2013. Enriching spatial data infrastructure (SDI) by user generated contents for transportation. *The International Archives of the Photogrammetry*. Remote Sensing and Spatial Information Sciences - ISPRS Archives XL-1/W3, 383–388.

Sharp, D., Anwar, M., Goodwin, S., Raven, R., Bartram, L., Kamruzzaman, L., 2022. A participatory approach for empowering community engagement in data governance: the Monash net zero precinct. Data & Policy 4, e5.

Singh, K.N., Behera, R.K., Mantri, J.K., 2019. Big data ecosystem: review on architectural evolution. In: Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing. Springer, Singapore, pp. 335–345.

Snyder, H., 2019. Literature review as a research methodology: an overview and guidelines. J. Bus. Res. 104, 333–339.

Solano, E.L., Catasús, M.G., Raffaghelli, J.E., 2023. (Open) data literacy: which relationships with open data adoption? A systematic review of the literature. *Italian Journal of Educational Technology*. Accepted Manuscript. Online. https://doi.org/10.17471/24994324/1303.

Song, S.H., Kim, T.D., 2013. In: A Study on the Open Platform Modeling for Linked Open Data Ecosystem in Public Sector. IEEE, pp. 730–734.

Stankovich, M., Behrens, E., Burchell, J., 2023. Toward meaningful transparency and accountability of AI algorithms in public service selivery. Dai's Center for Digital. Acceleration.

Styrin, E., Luna-Reyes, L.F., Harrison, T.M., 2017. Open data ecosystems: an international comparison. Transforming Government: People, Process and Policy 11 (1), 132–156.

Susha, I., van den Broek, T., van Veenstra, A.F., Linåker, J., 2023. An ecosystem perspective on developing data collaboratives for addressing societal issues: the role of conveners. Gov. Inf. Q. 40 (1), 101763.

Sveen, A.F., 2017. The open geospatial data ecosystem. KART OG PLAN 77, 108–120.

Tan, E. (2022). Designing an AI compatible open government data ecosystem for public governance. In *Proceedings of the 55th Hawaii International Conference on System Sciences* (pp. 2324-2333). ACM.

Tang, J., Zhu, H., Guo, J., 2022. Information diffusion between users in open data ecosystem: modelling and simulation analysis. Math. Probl. Eng. 2022 (2892274), 1–23.

Tron, C. (2020). Considering the web as a digital public space. In *Proceedings of the 3rd International Conference on Web Studies* (pp. 16-21). ACM.

Umbach, G., 2022. Statistical and data literacy in policy-making. Stat. J. IAOS 38 (2), 445–452.

Union, E., 2022. AI Watch European Landscape on the Use of Artificial Intelligence by the Public Sector. Publications Office of the European Union, Luxembourg.

Van den Homberg, M., Susha, I., 2018. Characterizing data ecosystems to support official statistics with open mapping data for reporting on sustainable development goals. ISPRS Int. J. Geo Inf. 7 (12), 456.

Van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F.J., Mulder, I., Alexopoulos, C., Flores, C.C., 2021. Towards value-creating and sustainable open data ecosystems: a comparative case study and a research agenda. JeDEM-eJournal of eDemocracy and Open Government 13 (2), 1–27.

Van Schalkwyk, F., Willmers, M., McNaughton, M., 2016. Viscous open data: the roles of intermediaries in an open data ecosystem. Inf. Technol. Dev. 22 (sup1), 68–83.

Vancauwenberghe, G., van Loenen, B., 2018. In: Exploring the Emergence of Open Spatial Data Infrastructures: Analysis of Recent Developments and Trends in Europe. Springer, Cham, pp. 23–45.

Vancauwenberghe, G., Dessers, E., Crompvoets, J., Vandenbroucke, D., 2014. Realizing data sharing: the role of spatial data infrastructures. In: Open Government. Springer, New York, NY, pp. 155–169.

Visser, M., 2013. Digital literacy and public policy through the library lens. Maine Policy Review 22 (1), 104–113.

Welle Donker, F., 2016. From access to re-use: a user's perspective on public sector information availability. *A+ BE| Architecture and the*. Built Environ. 21, 1–282.

Welle Donker, F., van Loenen, B., 2017. How to assess the success of the open data ecosystem? Int. J. Digital Earth 10 (3), 284–306.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3 (1), 1–9.

Wilkinson, M.D., Sansone, S.A., Méndez, E., David, R., Dennis, R., Hecker, D., Castro, L.J., 2022. Community-driven governance of FAIRness assessment: an open issue, an open discussion. Open Research Europe 2, 146.

Wilson, L., 2019. Understanding the data ecosystem. In *Data-driven Marketing Content*. Emerald Publishing Limited, Bingley, pp. 11–21.

Wilson, B., Cong, C., 2021. Beyond the supply side: use and impact of municipal open data in the US. Telematics Inform. 58, 101526.

Wolff, A., Wermelinger, M., Petre, M., 2019. Exploring design principles for data literacy activities to support children's inquiries from complex data. Int. J. Hum Comput Stud. 129, 41–54.

Xiao, N., & Miller, H. J. (2021). Cultivating urban big data. In *Urban Informatics* (pp. 547-565). Springer, Singapore.

Yunita, A., Santoso, H.B., Hasibuan, Z.A., 2022. 'Everything is data': towards one big data ecosystem using multiple sources of data on higher education in Indonesia. Journal of Big Data 9 (1), 1–22.

Zillner, S., Gomez, J.A., García Robles, A., Hahn, T., Le Bars, L., Petkovic, M., Curry, E., 2021. In: Data Economy 2.0: from Big Data Value to AI Value and a European Data Space. Springer International Publishing, Cham, pp. 379–399.

Zuiderwijk, A., Janssen, M., Van De Kaa, G., Poulis, K., 2016. The wicked problem of commercial value creation in open data ecosystems: policy guidelines for governments. Information Polity 21 (3), 223–236.