Information Systems and e-Business Management (2026) 24:657–695
https://doi.org/10.1007/s10257-026-00719-8

ORIGINAL ARTICLE

A conceptual deep dive into data ecosystems: a systematic literature review

Maarten de Mildt[1,2] · Sander Van Damme[1,2] · Sofie Verbrugge[1,2] · Didier Colle[1,2]

Received: 2 May 2025 / Revised: 18 December 2025 / Accepted: 5 March 2026 /
Published online: 16 March 2026
© The Author(s) 2026

Abstract

Data sharing in data ecosystems has gained interest as a means for innovation in industry, academia and governments. While there is an abundance of literature on data ecosystems and seminal works such as Oliveira et al. (Knowl Inf Syst 61(2):589–630, 2019) (pre-2017) offer foundational definitions of data ecosystems, recent studies still state that these remain unclear, complicating discussion and analysis of data ecosystems. Additionally, developments like the Common European Data Spaces and the International Data Space Association, which result from the European Strategy for Data, demand an updated study. This study bridges this gap through a systematic literature review on how data ecosystems have been defined in literature since 2017. We find that (1) while a generally accepted definition exists, being ”Data Ecosystems consist of a loose set of interacting actors that directly or indirectly consume, produce, or provide data and other related resources”, the broadness of this definition requires careful consideration in its application. (2) Data ecosystems originate from platform and ecosystem theory, each offering a different perspective on its functioning, contributing to contextual operationalization of data ecosystems. This results in six key concepts of data ecosystems: interconnection and interdependence, data value network, ecosystem metaphor, sociotechnical, networks of human and non-human actors, self-organizing. (3) The term data ecosystem is susceptible to trends and used to repackage established concepts as innovative. Understanding these perspectives is important to ensure the success of data ecosystems in practice. This nuanced approach highlights the importance of contextualizing data ecosystems, rather than using a rigid definition.

Keywords Data ecosystem · Data sharing · Data governance · Data space

Extended author information available on the last page of the article

1 Introduction

Data play an increasingly important economic and social role, driven by the proliferation of digital technologies which, in turn, have accelerated the growth of digital platforms, digital ecosystems and data ecosystems (Acs et al. 2021). Data sharing between organizations has proven to be a key driver of innovation (Zuiderwijk et al. 2014). Data ecosystems enable such data sharing by building collaborative networks while safeguarding trust, interoperability and sovereignty (Otto et al. 2022). The interest in data ecosystems is reflected in the European Strategy for Data, which aims to increase the availability of data by supporting data ecosystems (European Commission 2024). In response, initiatives like the International Data Space Association (IDSA 2024) propose common standards to enable data exchange. Initially highlighted by Rufus Pollock (2011), the term "data ecosystem" has since become central to discussions about data governance and data sharing, in the context of business platforms and ecosystems. Despite its prominence, it is argued in the literature that the definition and theoretical development of data ecosystems remain debated among scholars (Runeson et al. 2021; Howard and Hyland Wood 2024). While previous studies have systematically mapped the conceptualization of data ecosystems as "Data Ecosystems consist of a loose set of interacting actors that directly or indirectly consume, produce, or provide data and other related resources (e.g., software, services, and infrastructure)" (Oliveira et al. 2019:604), recent studies suggests that these analyses have not kept pace with emerging data sharing initiatives such as the Common European Data Spaces (EU 2024; Farrell et al. 2023) and the International Data Space Association (IDSA 2024). For example, the IDSA was founded later than the systematic review of Oliveira et al. (2019) and coined the term "Data Spaces". Additionally, as part of the European Strategy for Data (European Commission 2024), several data policy regulations have been put in place since, like the Data Governance Act (DGA) (Parliament and Commission of the European Union 2022) and the Data Act (DA) (Parliament and Commission of the European Union 2023). This raises the question of whether the increased uptake of data ecosystems has resolved earlier conceptual ambiguities. In addition, this signifies a need to better understand how data ecosystems are theoretically integrated within the broader, existing, concepts of digital platforms and digital ecosystems. Such an investigation must identify success factors that influence the efficacy of data ecosystems and, by extension, the governance and business models that support them. In essence, a lack of conceptual clarity regarding the "data ecosystem" concept continues to hinder effective communication and development.

A practical issue with this lack of clarity is that systems, networks and platforms get labelled as a data ecosystem all too often. The innovation benefits typically attributed to data ecosystems consequently are forced upon those systems, obfuscating the actual value they deliver (de Mildt et al. 2025). In the work by de Mildt et al. (2025) several use cases are discussed that are all given the label of a data ecosystem by the initiators of the cases. This shapes our view on a data ecosystem use case as a combination of a data sharing architecture/system, and a purpose ("use") for data that is shared through that system. Consider an illustrative example of a diploma sharing case where government databases are made accessible through an intermediary party

that allows citizens to consent to sharing their diploma to private companies. Such a case thus fits the intermediary view on data sharing of Möller et al. (2024). At the same time, initiatives like Catena-X, where data is shared along automative supply chains in use cases like traceability or carbon footprint reporting, are also given the label of a data ecosystem. Catena-X is described by Möller et al. (2024) as a Data Space, rather than an interaction driven purely by an intermediary. Thus, employing different conceptualizations of data ecosystems, and their use cases. As both are given the data ecosystem label in practice, both are assigned the same expected benefits, that align with the expected benefits of policy frameworks like the European Strategy for Data. This leads to such systems being considered for public funding and backing by policy makers. However, as mentioned, some expected benefits are more often than not, not reached, resulting in public funding being misallocated, and the actual goal to innovate of these policies to never be reached. Even more so, systems that deliver tangible value but adopt a more pragmatic stance regarding the limitations of current implementations may struggle to secure funding. Misusing the term data ecosystem, and overclaiming benefits associated with the term will thus lead to overconfident systems dominating the discourse of these policies.

Therefore, and since the concept continues to gain popularity, we aim to investigate if data ecosystem literature has reached a stage of conceptual clarity to avoid initiatives to fail, and funding to be misallocated. To resolve these issues, this research aims to (1) investigate how data ecosystems are conceptualized in literature and whether there is a common definition, (2) identify the key concepts from which these literature definitions draw and how they co-shape the concept of data ecosystems, and (3) evaluate whether these definitions can be applied to concrete use cases in a meaningful way. Here, applying them in a meaningful way means the definition should allow forming included and excluded (groups of use) cases, specifically to evaluate what achievable benefits in these cases are. For example, can it be determined if the exclusion group includes digital platforms that are not considered data ecosystems. In practice, we notice Data Space initiative participants struggle to apply and agree on (legal) definitions when setting up data sharing systems (Reiberg et al. 2022). While "soft definitions" are deliberately used to allow a variety of "forms of data spaces" in differing contexts to fit the definition (Hutterer et al. 2023; Frigeri et al. 2024; DSSC 2024), this does not mitigate the difficulty of assessing if cases will achieve the expected benefits of the European Strategy for Data. While context matters, it has not been addressed by providing an overview of what use cases different context and "soft definitions" result in.

The paper proceeds as follows. The methodology employed of the literature structure, which databases were queried, what the search strategy entailed and how results were coded, is described in Sect. 2. Next, the different results of Sect. 3 lay out the findings on the common data ecosystem definition in Sect. 3.1, six key concepts of those definitions in Sect. 3.2, data ecosystem use cases and trends present in data ecosystem studies in Sect. 3.3, and finally how data ecosystems position themselves relative to existing theory is described in Sect. 3.4. In Sect. 4 the main findings are discussed in light of their theoretical and practical/policy implications. The paper is naturally concluded in Sect. 5.

2 Related work

Research on data ecosystems and its conceptualization is grounded in related concepts. Accordingly, a brief description of the most relevant concepts is appropriate. Our understanding of the current state of data ecosystem conceptualization research, data sharing, ecosystem theory, and platform theory is presented. Given this work's focus on the applicability of theory and conceptualization, "explanation" theory type is considered as an appropriate perspective on information systems theory (Gregor 2006). We acknowledge that other perspectives on theory exist, and do not condone any of them, rather the "explanation" perspective fits our research goals of figuring out the status on what is, how, why, when, and where of data ecosystems. Where the why, when and where relate closely to the research goal of meaningful applicability of the "what is" of data ecosystems.

2.1 Data sharing

Data sharing is defined by Jussen et al. (2023:3688) as "Data sharing is the domain-independent process of giving third parties access to the data sets of others. These third parties may be other companies (usually not direct competitors), individuals, or public institutions. The shared data is often used to develop new applications and services. The expectation is to be compensated financially or through other benefits (e.g. receiving data) for providing the data. What the data may be used for and how it is made available is determined within the framework of the (legal) agreements between the data providers, data consumers, and other roles, depending on the use case". An important nuance is made by them, that data exchange is distinct from data sharing. Data exchange refers solely to the technical processes needed to share data. Data sharing also incorporates the organizational aspects of sharing data based on governance like usage policies and business model consequences that stakeholders will consider before sharing data. Data ecosystems are named to be a product of frequent data sharing between actors that take on different roles (Jussen et al. 2024). Data sharing is thus considered a fundamental practice in data ecosystems (Fassnacht et al. 2024). Remarkably, from a data sharing perspective, concepts like data spaces and data marketplaces are explicitly named next to data ecosystems (Fassnacht et al. 2024). While throughout the work they are often named in unison, sharing many related concepts and issues, this indicates they do not seem to explicitly consider data spaces a subset of data ecosystems or vice versa.

2.2 Data spaces

There seems to be a consensus among scholars on certain elements used to define data spaces, like the fact that they enable data sharing among different organizations, as illustrated by the following definitions. "Data spaces are decentralized data infrastructures designed to enable data-sharing scenarios across organizational boundaries by implementing mechanisms for secure and trustworthy data sharing—such as distributed data storage and the sharing of meta-data. They guarantee data sovereignty by ensuring that the data provider determines control over the access and use of the

shared data." (Möller et al. 2024:6). An example of this partial consensus can be found by the following definition that builds on this view: "Ecosystems building on a decentralized form of digital platforms, referred to as data spaces, […] Data spaces provide a secure infrastructure for interorganizational data sharing without requiring centralized integration. Instead, ecosystem participants maintain data sovereignty," (Ammann and Hess 2025:1). However, when it comes to positioning data ecosystems in relation to data spaces, scholars diverge on data space conceptualization. Fassnacht et al. (2024:47), while building on similar foundations as Möller et al. (2024) does, like Otto and Jarke (2019), explicitly distinguish between data spaces and data ecosystems by implying that the latter is a broader concept. In contrast Möller et al. (2024:41) conceptualize them as embedded in data ecosystems: "Data spaces allow the formation of flexible organizational forms that grant a delimited set of members access to a secure and trusted space to share data, which can be embedded in a larger data ecosystem.". Even within the literature stream that conceptualizes data spaces as (part of) an ecosystem, divergence exists. Curry et al. (2019)'s data space conceptualization is interpreted by Kari et al. (2024:177) as follows: "Data spaces are open platform ecosystems managed by a legitimized orchestrator that institutionalizes an alliance of different organizations". Kari et al. (2024) thus build on this view, arguing that while they also consider data spaces an ecosystem, they classify them as platform ecosystems rather than data ecosystems. Concluding this section, literature seems to show no consensus on whether a data space is synonymous with a (type of) data ecosystem or only a part of a data ecosystem.

This comes together in the mentioned "soft definition" data spaces suffer from, causing the data space concept to get redefined by studies to fit a purpose. As per Guggenberger et al. (2025:822)'s definition stating data spaces are "organizations compromising multiple legally autonomous entities […] with the shared goal of trustful decentralized data sharing", which avoids using any existing platform or ecosystem concept. This fitting for a purpose raises the question of whether uptake of the data space concept resolves conceptual ambiguity in its positioning relative to data ecosystems.

2.3 Platforms, ecosystems and platform ecosystems

As apparent in the previous section platform and ecosystem concepts define and conceptualize data ecosystems like in taxonomies and typologies that mention platform-centric organization, competition and cooperation, and decentralized control (Gelhaar et al. 2021; D'hauwer and Vandercruysse 2025; Guggenberger et al. 2020). Accordingly, an overview of what is considered platform and ecosystem theory here is appropriate. The main elements of platform theory concern platform economy and digital platforms in information systems. We consider platform economy as; "The platform economy comprises a distinctly new set of economic relations that depend on the Internet, computation, and data." (Kenney and Zysman 2016:66), emphasizing the dependency on data. Regarding digital platforms, an essential view is the market-based perspective, where a platform is a means to facilitate network externalities, also referred to as network effects (Hein et al. 2020). From a technical perspective a digital platform is typically considered an extensible codebase to which

complementary third-party modules can be added (De Reuver et al. 2018; Hein et al. 2020), emphasizing power of the platform as the central codebase, while making collaboration through third party software possible. This collaboration also led to a sociotechnical view on platforms as technical elements (of software and hardware) and associated organizational processes and standards (De Reuver et al. 2018). In the same work of De Reuver et al. (2018), the association between ecosystems and platforms is made, where ecosystems are defined as "A collection of complements (apps) to the core technical platform, mostly supplied by third-party" from a technical view, and as "Collection of firms interacting with a contribution to the complements." from an organizational perspective.

Each platform creates an ecosystem where value is created (Kenney and Zysman 2016). It is therefore not surprising that in recent years the term platform ecosystems emerged, where there is a core platform like an application store on which third party developers can sell applications, essentially coordinating an ecosystem of supply and demand for the core digital platform (Parker et al. 2016; Hein et al. 2020). A similar view that emphasizes the centrality of the platform in a platform ecosystem can be found in Schreieck et al. (2022) who refer to platform ecosystems as "a set of companies that co-create value on a digital platform […] Typically, one company owns the digital platform and acts as ecosystem orchestrator […]". Platform ecosystems thus seem to represent a middle-ground between the ecosystem idea of collaboration and cooperation and the control of a central platform that remains. However, as with data ecosystems, the platform ecosystem concept can be critiqued, as general platform theory also considers users to create value through network effects through a central platform that orchestrates interactions.

3 Methodology

Given these research aims, we employ a systematic literature review methodology aimed at examining how data ecosystems are defined, and how they are linked to existing theoretical frameworks. We base our study on (Attard et al. 2015), and (Oliveira et al. 2019) as their focus lies with synthesizing the conceptualization of data ecosystems in academic literature. The procedure is as follows: define a search strategy, define sources and selection criteria, and perform the coding of selected sources.

3.1 Search strategy

We conducted a literature search on the ISI Web of Knowledge (which includes Springer articles), SCOPUS, ScienceDirect, ACM, and IEEE databases using the keywords "data ecosystem" or "data ecosystems" and filtered on the abstract and title contents. The resulting number of studies for each database queried are shown in Table 1. Here we deviate slightly from the strategy employed by Oliveira et al. (2019) in search terms and databases. Regarding search terms, next to "data ecosystems", they included terms like "Big Data Ecosystems", however these are logically included in "data ecosystems" and thus deliver no additional results. We surveyed the

Table 1 Searched databases and the number of studies extracted from them by the "data ecosystem[s]" query

Database	nr. of Studies
ACM	50
IEEE	19
ScienceDirect	762
SCOPUS	387
Web of Knowledge	326

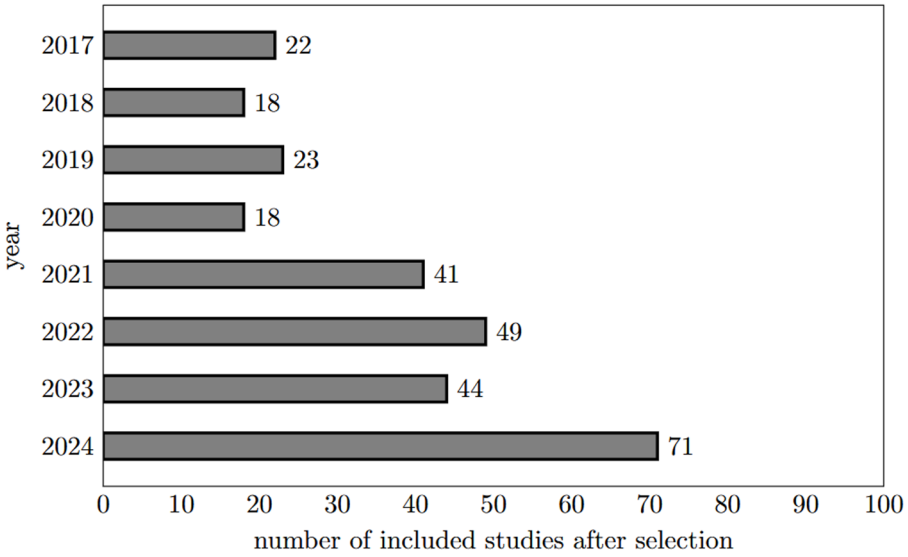


Fig. 1 Number of studies published on data ecosystems from 2017 onward, until 2024 after filtering to ensure studies were about data ecosystems. The increasing importance of data ecosystem studies continues past 2017 when the study by S. Oliveira et al. (2019) was conducted and does not seem to have reached a halt just yet. This is reason to believe an update is appropriate

same databases as Oliveira, with an expansion to include all of Web of Knowledge (which includes Springer, as queried by Oliveira's study). As they already investigated results prior to 2017, we limited our results from 2017 up until 2024. Two other reasons for this timeframe are the increase in data ecosystem initiatives and policies which came into fruition since Oliveira's study (see the introduction), and an increase in academic literature published from 2017 onward, as shown in Fig. 1.

The relevance of each article was initially assessed based on the title and abstract. If both the title and abstract made clear that the study was not about "data ecosystems" or related concepts (e.g. data sharing, data science…) it was filtered out in this first iteration. For example, several articles were filtered from environmental science studies that featured the word "ecosystem", and the word "data" from describing data gathered for their research, but did not mention 'data ecosystem' as such. If the relevance remained unclear after reading the title and abstract, we read the full article in similar fashion, checking if the study is indeed on data ecosystems. For both iterations, the formal inclusion and exclusion criteria were based on those defined by Oliveira et al. (2019), as follows:

Inclusion criteria

- The work focuses on data sharing from an organizational, business, or legal perspective.
- The work positions data ecosystems as a model for data sharing.

Exclusion criteria.

- The work does not directly or indirectly define the term "data ecosystem".
- The work describes a purely technological data (pipeline) solution.
- The work describes single-organization solutions that do not involve data sharing across organizations.
- The work is not a journal article.
- The exclusion criteria defined by S. Oliveira et al. (2019) and not already mentioned above:
  - The study is not peer-reviewed.
  - The study contains less than one page.
  - The study is not written in English (or does not have an original translation).
  - The study is not accessible on the Web.
  - The study is duplicated.

Figure 2 shows the filtering process, including the number of studies (records) filtered in each step. The filtering process resulted in 286 selected studies, starting from 1544 hits from the combined databases, as depicted in Table 1. From the initial set of 1544 records, 377 records were filtered as duplicates found in multiple databases. 1167 Were screened, in first iteration on title and abstract, and in second iteration on full paper content, of which 716 were filtered on the premise of not dealing with data ecosystems, and 165 additional ones because they featured no definition of a data ecosystem.

3.2 Coding of selected sources

We employed a three-phase coding system (open, axial and selective) (Williams and Moser 2019) to identify how data ecosystems are conceptualized in data ecosystem literature. The purpose of open coding is to break down the studies into minimal parts, the axial coding step refers to linking these and establishing overarching categories, and the selective coding step refers to linking these categories to an overarching theoretical framework. The coding was performed by two of the authors,

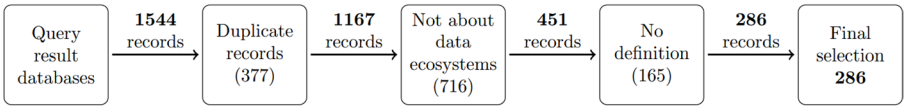


Fig. 2 The filtering of the combined records from the queried databases, resulting in 286 selected studies

to consider different viewpoints as much as possible, as, one coder had a business and technical background, and one coder had a social and governance background. To resolve conflicts frequent discussions were held between both coders to ensure a continuous internal feedback loop. In such discussions, iterating over the coding and providing argumentation as to why a certain code was considered applicable was crucial to decide in conflict. If this did not resolve conflict, an additional discussion was held with the two remaining authors. Finally, other experts in the field of data ecosystems (social, economic, and legal background) were consulted twice where we presented intermediate results and gathered additional feedback to ensure validity of the coding. To ensure additional intercoder reliability, a first round of coding several studies (but not yet the whole set) was performed by both coders and then revised together and adapted where needed. This way, a common understanding was established of relevant codes and types of excerpts to select, as well as ambiguity in the meaning of terms clarified.

Two major frameworks, platform and ecosystem theory, were identified (cf. Fig. 4), which were uncovered in two follow-up coding iterations. Codes were linked to their respective theories by checking if the references used in the studies referred to platform or ecosystem literature, or, in ideal cases, directly mentioned platforms or ecosystems. In case of no direct link between a statement and literature, the positioning in literature of the study by the related work was used if there was one. If neither were present, we connected this to data ecosystem theory (as naturally, all selected studies were positioned in data ecosystem literature). It must be noted that multiple studies connected to multiple theories, so they are not mutually exclusive. In addition, we included literature that was referenced in the data ecosystem, platform and ecosystem definitions in the selected papers. This way, we aimed to better understand the origin and context of these frameworks in relation to data ecosystems. The coding resulted in 639 open, 145 axial, and 22 selective codes.

The three-phase coding performed is further explained by an example excerpt from one of the included studies. The following quote from Susha et al. (2022) *"[...] a loose set of interacting 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 by collaborating and competing with each other promote data ecosystems [...] Heimstädt, Saunderson, and Heath (2014) further point out the circular flow of resources and dependencies between suppliers, intermediaries, and users as properties of data ecosystems."* was open coded as **"loose group of actors"**, **"interacting actors"**, **"actor collaboration and competition"**, **"consume, produce, or provide data "**, **"circular flow of resources"** and **"circular dependencies between suppliers, intermediaries and users"**. Each of these were extracted because they highlight a certain characteristic that this study attributes to data ecosystems. Such characteristic can be either a concept (e.g. "loose group of actors"), an action (e.g. "consume, produce, or provide data"), or a boundary resource (e.g. "actor collaboration and competition"). These open codes were then clustered into axial codes in the second step. For instance, "loose set of interacting actors" and "varying interactions and circular interdependencies" were coded as **"loosely coupled"** since they both concern the relation between actors, while "consume, produce, or pro-

vide data" and "circular flow of resources" was coded as **"facilitating data interactions"** since they both describe how resources like data are used by actors. In a third, selective coding step, these overarching clusters were then further bundled into the top level codes, for instance, "loosely coupled", "facilitating data interactions" and "interconnected network" are coded as **"interconnection and interdependence"** as they all relate to actor interactions and relations. Figure 3 denotes a sample of this larger quote to visualize how each study was coded.

4 Results

The results of the systematic mapping are discussed in this section. First, we explain how there is in fact a common, established, definition of data ecosystems in Sect. 4.1. However, it remains abstract and difficult to apply to tangible use cases. Second, six key concepts, originating from platform and ecosystem theory, which make up the main characteristics of a data ecosystem were identified in Sect. 4.2. We present three categories of data ecosystem use cases and common terms to describe these use cases in Sect. 4.3. In addition, we noticed an apparent division in the analysis of use cases. These use cases can be categorized according to two dimensions (actor diversity, and platform centrality) to provide further insights into data ecosystem functioning, as discussed in Sect. 4.4.

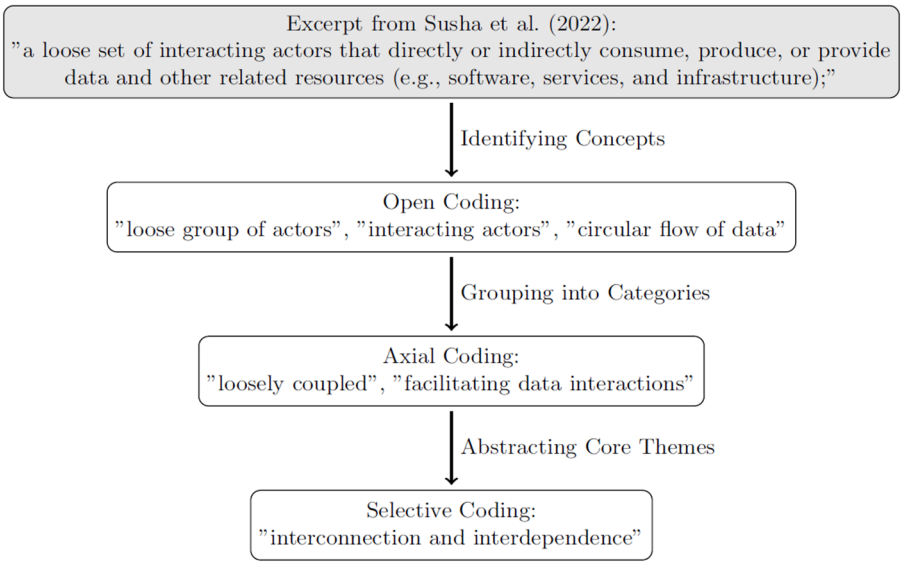


Fig. 3 The coding procedure employed, starting from extracted literal quotes, moving through the three-phase coding steps to result in a limited set of central concepts described in the studies

4.1 Established data ecosystem definition

A recurring theme in data ecosystem studies, is that there is no generally accepted definition among scholars (Oliveira et al. 2019; Shah et al. 2021; Runeson et al. 2021). This problem is mitigated in two ways: either a new definition is proposed based on existing definitions, or a set of existing definitions is considered. However, since Oliveira et al. (2019) published their study, it has been by far the most common reference used to define a data ecosystem. Oliveira et al. (2019), in turn, base their definition on, among others, commonly cited works of Zuiderwijk et al. (2014) and Dawes et al. (2016). The set of commonly cited studies that refer to Oliveira' study is as follows: (Linåker and Runeson 2020; Runeson et al. 2021; Susha et al. 2022; Reggi and Dawes 2022; Gupta et al. 2020; Shah et al. 2021). Showing that highly cited works in data ecosystem literature cite Oliveira's work, further spreading the definition. Here, "commonly cited" refers to being commonly cited to define data ecosystems in the study selection of 286 papers. Therefore, we can state the following:

1. **The definition put forward by** Oliveira et al. (2019) **is generally accepted**: "Data Ecosystems consist of a loose set of interacting actors that directly or indirectly consume, produce, or provide data and other related resources (e.g. software, services, and infrastructure)".
2. While Oliveira's definition gets as close to a consensus as feasible, the underlying critique persists that **what is**, **and is not**, **a data ecosystem cannot easily be determined via this definition.** Several open questions remain concerning which type of use cases and popular technologies fit the data ecosystem category.

4.2 Six key concepts of data ecosystems

The coding of the selected literature revealed six key concepts defining data ecosystems (see Fig. 4). These concepts, rooted in both platform and ecosystem theory, are interconnection and independence, data value network, ecosystem metaphor, sociotechnical, networks of human and non-human actors, self-organizing. In the following sections, we discuss each of the key concepts and discuss how they relate to platform and/or ecosystem theory. When discussing ecosystem or platform theory fundamentals, we refer to the related work section to provide background.

4.2.1 Interconnection and interdependence

In ecosystems theory, one of the fundamental principles is economic cooperation and coexistence. Literature building on these principles conceptualizes actors within data ecosystems as interconnected and interdependent (Howard and Hyland-Wood 2024; Supinajaroen et al. 2022; Ghazinoory et al. 2021). Each actor (whether an organization, individual or technology) interacts with others, which creates a web of relationships. These interactions and the roles actors play in them evolve over time based on feedback mechanisms within the ecosystem. With respect to data ecosystems, this is linked to the need for governance mechanisms, given the varying interests and types

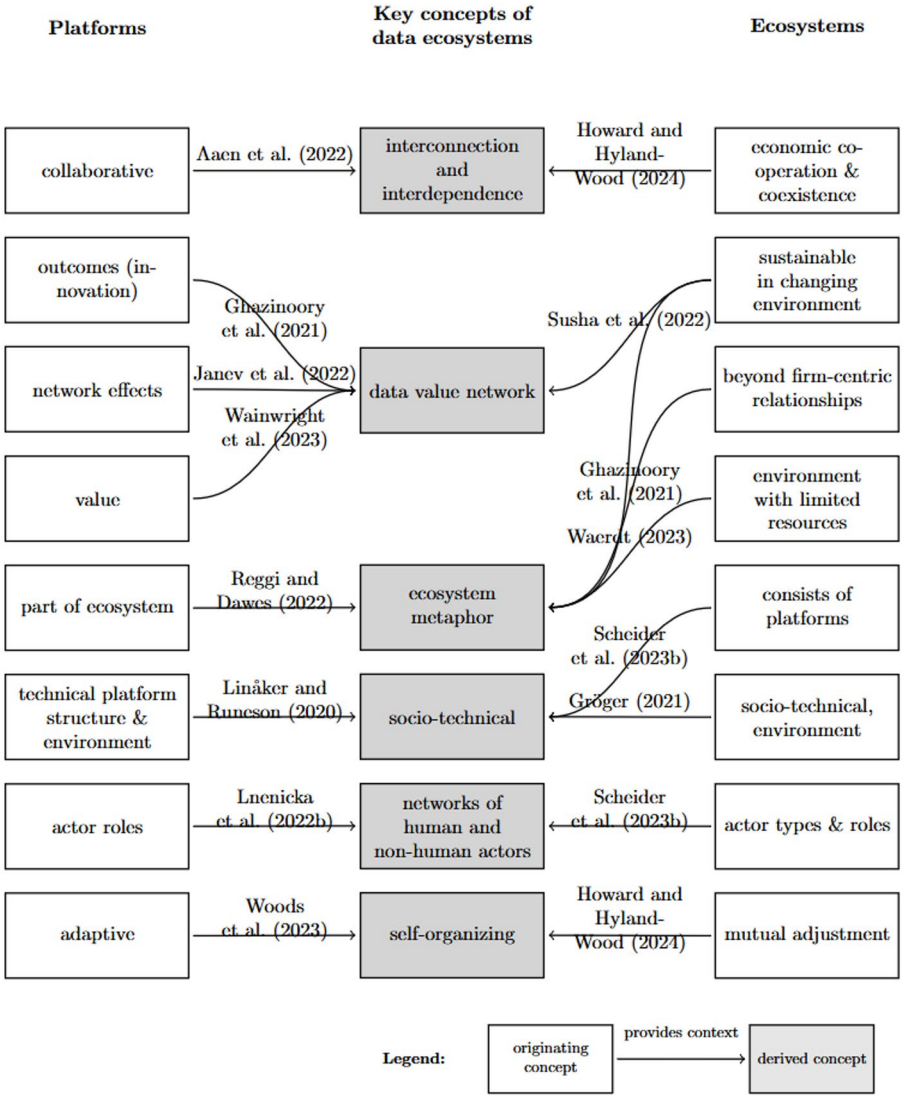


Fig. 4 Coding results of data ecosystem definitions, revealing six essential concepts that define data ecosystems. Several example relevant studies that provide context to their meaning are given for each concept. Originating from platform and ecosystem theory, each code is given additional meaning depending on these two perspectives. This not only reveals that data ecosystem definitions originate from these two theories but also demonstrates that the perspective taken is essential

of actors. Essentially, adding to the data ecosystem concept that (the roles) actors co-evolve through self-reinforcing and as a response to innovation. Such governance mechanisms must safeguard shared visions of expected benefits for the ecosystem. Essentially emphasizing that giving data and knowledge back to the ecosystem to provide value for all requires cooperation, even among competitors with different pursued values.

From a platform theory perspective, platforms are collaborative systems (Aaen et al. 2022; Woods et al. 2023; Feyzbakhsh et al. 2022). They connect different types of users, with the platform functioning as a collaborative network mechanism. This causes organizational boundaries to disperse and value to be created by connecting different types of users (Wainwright et al. 2023), and additional users mean additional data points, which creates network effects and thus increases the value of the platform (Lee et al. 2020; Woods et al. 2023). The emphasis on value creation through collaboration among actors via the platform leads to a similar conclusion as ecosystem theory: actors are mutually dependent for value creation. Accordingly, typical conceptualization of data ecosystems coming from both platform and ecosystem theory leads to interdependent and interconnected actors.

4.2.2 Data value networks

Platforms form the interface between data and new forms of value creation (Woods et al. 2023). Combining this with the ecosystem element of the changing environment to which this value creation must adapt, value networks are used to conceptualize data ecosystems (Ghazinoory et al. 2021; Runeson et al. 2021; Janev et al. 2022; Lnenicka et al. 2022a, b; Woods et al. 2023; Wainwright et al. 2023). In this context, data is a key resource for creating value, as it's cyclical nature in data ecosystems is what turns traditional value chains into value networks. Essentially, data of many types is subject to feedback loops which eventually returns transformed data to the ecosystem. Making data an actionable good through these feedback loops, is believed to bring innovation and knowledge sharing.

Regarding the ecosystem theory side to data value networks, distinction between a network and an ecosystem is made. In a network, the interconnections between actors do not form an existential threat, whereas in an ecosystem they do. As such, in a data ecosystem, this value network should remain sustainable in changing environments, both from internal and external changes (Susha et al. 2022; Calzati and Loenen 2023; Corbett et al. 2020). While value is a generic concept that is present in both platform and ecosystem theory, value is a result of platforms within the ecosystem.

4.2.3 Ecosystem metaphor

Regarding the changing environment, papers touching upon general ecosystem theory systematically referred to the ecosystem metaphor (and similar terms like digital business ecosystems, and innovation ecosystems) to conceptualize a data ecosystem, drawing inspiration from the biological counterpart (Corbett et al. 2020; Diran et al. 2020; Ghazinoory et al. 2021; Feyzbakhsh et al. 2022; Susha et al. 2022; Calzati and Loenen 2023; Woods et al. 2023; Waerdt 2023; Howard and Hyland-Wood 2024; Shah et al. 2024). This means that data ecosystems must be sustainable in changing environments with limited resources, like funds for innovation or knowledgeable people, similar to a natural ecosystem where resources (or the lack thereof) dictate balance. Changes to the environment come from the beyond firm-centric relationships that are essential to the ecosystem view, even crossing industries (as the biological counterpart traverses species). Describing data ecosystems, as an analogy to their

biological counterpart, however, is somewhat contradictory given it describes it using a tautology. Using the ecosystem metaphor description describes a data ecosystem as an ecosystem. The ecosystem metaphor itself is referred to as "the relationships and interactions between living organisms and their environment" (Schulze et al. 2005; Briscoe and De Wilde 2006), showing it resembles Oliveira's definition as similar terms like interactions between actors in the ecosystem are used.

Many of the "living organisms" in this environment are platforms, as platforms are considered a crucial part of a data ecosystem (Calzada and Almirall 2020; Reggi and Dawes 2022; Molder et al. 2022). It is seen both as a technical underpinning, and an organizational structure to facilitate the network that a data ecosystem consists of. Data ecosystems experience network effects that are (in part) created by their underlying platforms, where the distribution of limited resources (e.g. partnerships with other platforms) play a role in each individual platform's success, as with individual species in the biological counterpart.

4.2.4 Sociotechnical

Many consider data ecosystems to be a sociotechnical system, including Oliveira et al. (2019), consisting of technical infrastructure and environmental context factors (Linåker and Runeson 2020; Gröger 2021; Jean-Quartier et al. 2022; Shadbolt et al. 2022; Lnenicka et al. 2022b; Supinajaroen et al. 2022; Scheider et al. 2023b; Howard and Hyland-Wood 2024). An important implication of a data ecosystem being more than just a technical implementation is that technology does not define a data ecosystem. This links back to both platform and ecosystem theory, given both theories consider technical, sociotechnical and organizational definitions. From the ecosystem perspective, the social, organizational environment consists of people and their capabilities (e.g. knowledge), as well as symbolic resources (e.g. branding and cultural perception) that influence the system. Here the distinction between a platform, often considered a technology, and an ecosystem, which is a system that goes beyond just the technical aspect, becomes clear. Combing the technical platform and its environment with the sociotechnical environment (which is typically conceptualized as consisting of platforms) leads to the socio-technical view being dominant in data ecosystem conceptualization.

4.2.5 Networks of humans and non-humans

Public or private individuals or organizations, platforms and machines can all make up a data ecosystem, as their conceptualization sets no boundaries on which actors it encompasses (Susha et al. 2022; Scheider et al. 2023b; Prinsloo et al. 2023). This links back to ecosystem theory where different types and roles of actors (or species, in the biological counterpart) are present. Many data ecosystem studies define or provide overviews of roles for these actors (like keystone/orchestrator, data providers, data consumers). However, there is no clear consensus on what each role encompasses, nor if there are roles that define data ecosystems.

From a platform perspective, platforms in data ecosystems are referred to as both technical and social entities that "collect, store, process, or act upon data" (Corbett et

al. 2020; Kitsios et al. 2021; Runeson et al. 2021; Gordon-Tapiero et al. 2022, p. 119; Lnenicka et al. 2022b). These include APIs, interfaces, portals and accommodating tools, leading to the non-human actors that are considered from this technical perspective. Therefore, a network of actors, human and machine, combine both platform and ecosystem understanding of actors. Although there is no consensus on what roles and types of actors exist and what the definition of each role or type is, there is consensus on the diversity of the actors in both platform and ecosystem theory.

4.2.6 Self-organizing

On an individual level each actor in a data ecosystem is autonomous, therefore, data ecosystems are commonly seen as self-organizing systems that can adapt to change due to this autonomy of its actors (Woods et al. 2023; Scheider et al. 2023b; Czvetkó and Abonyi 2023; Howard and Hyland-Wood 2024). A central idea is that actors mutually adjust, meaning that the decisions of one actor steer decisions of others. This adjustment happens while there is no central/hierarchical governance or control over the ecosystem, thus, each actor independently decides how to adjust to other actors' decisions. From a platform perspective, platforms are also referred to as adaptive and processual, as they are not predetermined in terms of how they are orchestrated (only in who orchestrates balance in the system, being the platform) and what their effects are (Woods et al. 2023). From an ecosystem perspective a very similar view is found. Mutual adjustment in an ecosystem refers to the ability of actors within an ecosystem to independently make decisions that consider the actions and needs of other actors. While related to interconnection and interdependence, which deals more with the value creation side, this self-organizing aspect is about adapting to organizational conditions and governance rules for actors to partake and survive in the system.

4.2.7 How key are these key concepts?

Having established what concepts are most used in literature to define data ecosystems, a brief discussion on the validity of these concepts to conceptualize data ecosystems is appropriate. Each of these concepts are used to try and distinguish data ecosystems from just any platform or ecosystem. However, each of these concepts can also be contested to define data ecosystem compared to platform and ecosystem theory. For example, the environmental counterpart ecosystem referred to as the ecosystem metaphor can be seen as an analogy for data ecosystems, rather than a defining concept. Another example is the interconnection and interdependence distinguishing data ecosystems from platform theory where the dependence is to platform(s), while for data ecosystems these dependencies are on individuals, intermediaries and many other possible actors. This point is also used to raise the idea for data ecosystems that these relations can evolve more quickly than the relative robustness of dominant platforms. However, at the same time the argument can be made that there is no inherent difference between interacting through a platform (after all platforms are considered part of an ecosystem) or an ecosystem. Platforms also connect a multitude of actors. An individual can still act independent from the platform, and as such the relation with that individual can evolve equally quickly with or without a dependency on a

platform. It is therefore challenging to attribute a weight to each concept that would describe how "key" they are to conceptualize data ecosystems.

4.3 Data ecosystem use cases

A third aspect that is used to define or at least provide an example of what a data ecosystem is, is use cases. In data ecosystem literature, analysis of use cases is common to extract or validate findings and show practical relevance. An overview of the different use cases found, categorized by their main theme, seems appropriate for obtaining an idea of what the literature considers instances of data ecosystems. It must be noted one study can include multiple case studies, of different types, and as such can be included in different categories. At the same time, not all studies investigate use cases, and as such not all 286 studies are represented here. The three distinct categories identified are: Open Data Ecosystems (ODE), Sector-based Ecosystems, and Technology-centered Ecosystems. Figure 5 provides an overview of the different use cases found: 51 ODEs, 84 Sector-based Ecosystems, and 64 Technology-centered Ecosystems. For ODEs, general open data initiatives (21), government driven initiatives (17) are the most popular. Sector Based ecosystems show a clear dominance of medical (33) use cases. The field is more nuanced for technology-centered ecosys-

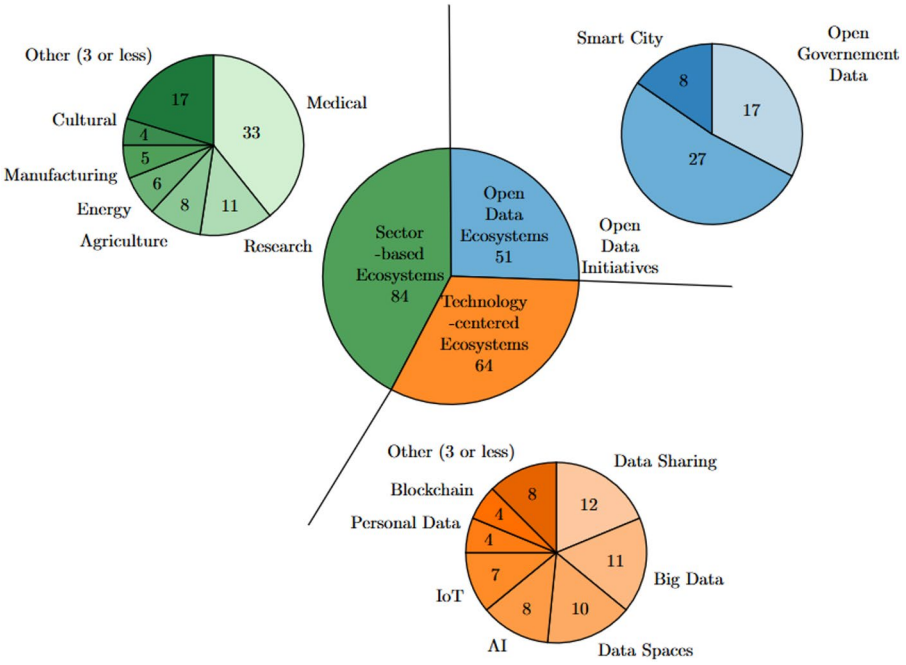


Fig. 5 An overview of the different use cases found, categorized by their main theme, is given. The three distinct categories found are: 51 Open Data Ecosystems, 84 Sector-based Ecosystems, and 64 Technology-centered Ecosystems. What this figure shows is a wide variety of use cases, and as such no clear boundary of what can or cannot be considered a data ecosystem

tems. Data Spaces (10), Data Sharing (12) and Big Data (11) technologies are among the more popular use cases.

4.3.1 Open data ecosystems

A first cluster is found in the ODEs, shown top-right in Fig. 5, which are data ecosystems that specifically focus on data sharing under public licenses (Runeson et al. 2021), also commonly referred to as open data. Open Government Data initiatives and platforms are common among ODEs. ODEs haven been a widespread phenomenon in recent times, ranging from studies performed in Europe (Linåker and Runeson 2020; Kitsios et al. 2021; Reggi and Dawes 2022; Ruijer et al. 2023; Lnenicka et al. 2024), Canada (Corbett et al. 2020), China (Tang et al. 2022), Iran (Feyzbakhsh et al. 2022) and several other open government data studies (Khurshid et al. 2019; Linåker and Runeson 2021; Boulet 2022; Nehra and Bansode 2023; Lnenicka et al. 2024; Shah et al. 2024; Fang et al. 2024; Fang and Zhao 2024). Given their prominence, many open data publishers remain government institutions. In addition, Several European Union member states are at the forefront of Open Government Data, being the dominant group within these ODE studies, and as mentioned in the EU's yearly Open Data Maturity report (EU 2024). Logically, such government initiatives are supported by public funding, emphasizing the importance of such funds being allocated to initiatives that achieve benefits attributed to data ecosystems.

Next to government open data, more general Open Data Initiatives are found that typically represent wider cooperation past only government open data. Examples include Research Open Data (Homberg and Susha 2018; Villarreal et al. 2024), Health Open Data (Begany and Martin 2017; Heijlen and Crompvoets 2021; Shadbolt et al. 2022; Vasto-Terrientes 2023), Cultural Open Data (Lee 2019; Booth et al. 2021), Open Data Observatories (Hamed et al. 2024), Environmental Open Data (Fenwick and Jurčys 2023), Agriculture (Chamorro-Padial et al. 2024), Spatial open data (Supinajaroen et al. 2022; Molder et al. 2022), Industry and Automotive (Runeson et al. 2021), Energy (Diran et al. 2020), Tourism (McLeod 2023), and several others (Styrin et al. 2017; Ward et al. 2018; Loenen et al. 2021; Lnenicka and Nikiforova 2021; Linåker and Runeson 2022; Lnenicka et al. 2022a; Skelly and Chiware 2022; Kamariotou and Kitsios 2022; Heltweg and Riehle 2023; Chen et al. 2024). A final common type of ODE is found in smart city open data initiatives (Aguilera et al. 2017; Barns 2018; Gupta et al. 2020; Lupi and Antonini 2020; Lnenicka et al. 2022b; Oldenburg and Pussinen 2024; Ma and Huang 2024; Wang et al. 2024). Even though there is not necessarily an explicit government party involved in these ODEs, public backing is common and the importance of its proper allocation remains.

When reflecting upon these use cases, they indicate that making data as the key resource available and allowing an ecosystem of autonomous actors to grow around it organically. Publicly making data available, also is commonly paired with a technical platform to make data exchange, finding data, and collaboration possible. However, most open data use cases do not conform with the view that platforms are just one part of the ecosystem, and that an ecosystem consists of platforms. These open data initiatives are typically instantiated as a single platform or orchestrating entity, from which the true network of platforms still needs to form.

4.3.2 Sector-based ecosystems

A second cluster of data ecosystem instances are those that have sector specific attributes like protocols, standards or technologies as part of the implementation. The cases in the top-left of Fig. 5 focus on sharing data with specific parties that operate in the same sector, for specific purposes. They differ from ODE in the sense that in sector-based ecosystems the type of actors that will make use of the data, and what they need it for, are known beforehand, like a medical professional getting access to aggregated patient data from multiple hospitals. Sector-based ecosystems, like Open Government Data, are a geographically widespread phenomenon. An abundance of sectors are covered by these ecosystems.

They include: Medical, which represents by far the biggest group of the sector-based ecosystems, that focus on making medical data available across practitioners and hospitals for diagnosis and research (Begany and Martin 2017; Casanovas et al. 2017; Grossman 2018; Thinyane 2018; Boyles et al. 2019; Geldof et al. 2019; Malthouse et al. 2019; Iacob and Simonelli 2020; Alfano et al. 2021; Donida et al. 2021; Geisler et al. 2021; Gout et al. 2021; Aaen et al. 2022; Baldacci et al. 2022; Ferretti and Vayena 2022; Pratt et al. 2022; Chu et al. 2024; Denecke et al. 2023; Flores-Toro et al. 2023; Hussein et al. 2023; Keim-Malpass and Kausch 2023; Mahmood et al. 2023; Näher et al. 2023; Sakor et al. 2023; Touré et al. 2023; Velmovitsky et al. 2023; Altendeitering et al. 2024; Cruz-Almeida et al. 2024; Ertürkmen et al. 2024; Howard and Hyland-Wood 2024; Palojoki et al. 2024; Pellizzoni and Falavigna 2024; Rojas-García et al. 2024), Research where data is shared among researcher institutes, stakeholders relevant for research, and possible participants (Geisler et al. 2021; McCord et al. 2021; McMahon et al. 2022; Fraga et al. 2022; Ramalli and Pernici 2023; Exner et al. 2023; Kemmer et al. 2023; Hawrylycz et al. 2023; Wendelborn et al. 2023; Hansen et al. 2024; Kyoda et al. 2024), Agriculture in which the focus is typically on making sure data flows smoothly along the value chain (e.g. from farmer to retailer) and environmental reporting (Eastwood et al. 2021; Sharma et al. 2022; Winters et al. 2022; Hrustek et al. 2022; Williams et al. 2023; Sestak and Copot 2023; Kapoor et al. 2024; Chamorro-Padial et al. 2024), Energy showing a similar need to agriculture as data needs to move between energy service providers and consumers alike (Immonen et al. 2020; Janev et al. 2021, 2022; Bao et al. 2022; Fu et al. 2024), Manufacturing focusing on supply chain data (Weber et al. 2022; Olson and Liu 2023; Pohlmeyer et al. 2024; Altendeitering et al. 2024; Li et al. 2024), Cultural that often aims to create digital collections with active participation from artists, curators, experts, government, art collectors (Lee 2019; Given et al. 2023; Liang et al. 2023; Sotirova-Valkova 2024), and several others (Allemang 2019; Knieps 2019; Scassa 2019; Rix et al. 2021; Möller et al. 2022; Pascalis 2022; van Donge et al. 2022; Chevallier et al. 2023; Legenvre and Hameri 2023; Liva et al. 2023; Wilkinson et al. 2023; Antolini et al. 2024; Konstantinidis et al. 2024; Kruk et al. 2024; Palviainen et al. 2024).

A reflection can again be made on these use cases in terms of how they relate to data ecosystem definitions as discussed. Economic cooperation and coexistence, the environment with limited resources, and the beyond firm-centric relationships are already established at the initiation of the initiative by the actors. The problem these actors solve is not establishing the ecosystem but establishing sustainable value cre-

ation. These actors are willing to cooperate but lack the means to do so. In most of these cases, a technical platform, standard, or protocol agreed upon by all actors is implemented to solve this problem.

4.3.3 Technology-centered ecosystems

A third use case cluster at the bottom of Fig. 5 is grouped on a more fundamental level of technological (concept) choice. They differ from the other two use case categories in the sense that these cases do not focus on a type of data, or type of actor, but rather on the tools needed to share data. The assumption is that if such tools exist, additional use cases will emerge in the ecosystem.

The different technologies and concepts that literature studies are: Big Data (Nadal et al. 2017; Hidalgo et al. 2018; Geldof et al. 2019; Khurshid et al. 2019; Koren et al. 2019; Côrte-Real et al. 2020; Francia et al. 2021; Ye et al. 2023; Narayanan et al. 2024; Shah et al. 2024; Taherdoost 2024), AI (Blaiszik et al. 2019; Gröger 2021; Susha et al. 2022; Prinsloo et al. 2023; Toorajipour et al. 2024; Nelson et al. 2024; Li et al. 2024; Nam et al. 2024), Personal data management platforms (Ottaviano et al. 2019; Calzada and Almirall 2020; Vandercruysse et al. 2023; Xu et al. 2023), Data Sharing (Iacob and Simonelli 2020; Hayashi et al. 2021; Kiesel et al. 2021; Runeson et al. 2021; Susha et al. 2022; van Donge et al. 2022; Chevallier et al. 2023; Legenvre and Hameri 2023; Rong and Luo 2023; Scheider et al. 2023a, b; Aho 2024; Kyoda et al. 2024), Data Spaces (Janev et al. 2021; Jean-Quartier et al. 2022; Czvetkó and Abonyi 2023; Legenvre and Hameri 2023; Woods et al. 2023; Möller et al. 2024; Noardo et al. 2024; Palojoki et al. 2024; Pohlmeyer et al. 2024; Sotirova-Valkova 2024), Internet of Things (Karanja et al. 2018; Côrte-Real et al. 2020; Runeson et al. 2021; Hyysalo et al. 2022; Kumar et al. 2023; Liang et al. 2023; Toorajipour et al. 2024), Blockchain (Khezr et al. 2022; Seneviratne et al. 2023; Xu et al. 2023; Lu et al. 2024), and several others (Curry and Sheth 2018; Linåker and Runeson 2020; Khezr et al. 2022; Gordon-Tapiero et al. 2022; Grassi and Lanfranchi 2022; Rajagopal 2023; Waerdt 2023; Taherdoost 2024).

These use cases show that a technical concept can also be used to initiate a data ecosystem. Here, the concept of a technical platform and environment in the context of data ecosystems is given more meaning. Not just platforms that allow interaction for data exchange, but also standards and general technological concepts, like Open Source, are seen as platforms in the ecosystem by scholars.

4.3.4 Data ecosystem trends

Next to use cases, data ecosystems are described by research fields that include specific sectors or technologies. For example, terms like ”Big Data Ecosystems” appear to describe a data ecosystem (e.g. ”government big data ecosystem” (Shah et al. 2024), or simply ”Big Data ecosystems” (Nadal et al. 2017). We mapped different terms used to discover trends in data ecosystem research over the years in Fig. 6. Some terms have lost popularity over the years (Big Data Ecosystems were most popular around 2019), while others have only recently gained popularity (Data Spaces only being increasingly used since 2022). Essentially, showing that the con-

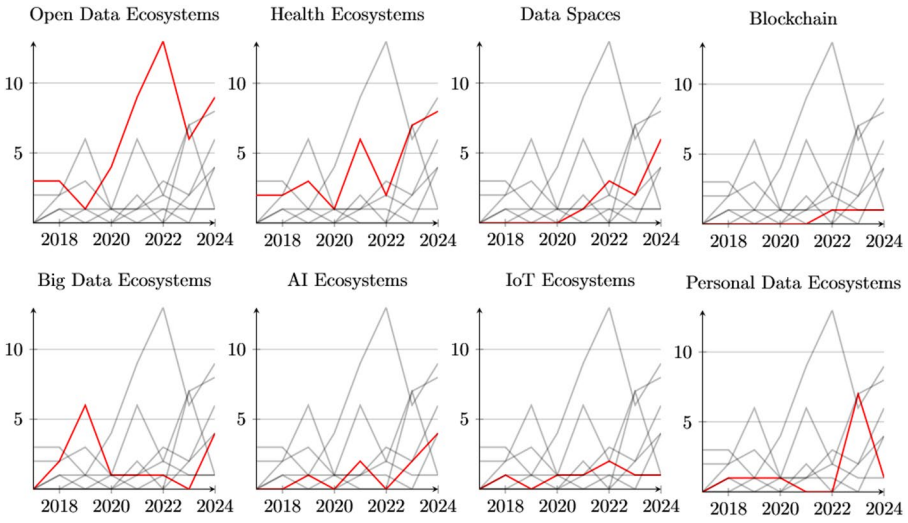


Fig. 6 Terms/research fields used to describe data ecosystems in which data ecosystems studies were found mapped over the years for which our study was conducted (2017–2024). For each sub-graph, the highlighted (red) line shows the mapping for the concept shown on top of the sub-graph. These show that the concept of data ecosystems is used to describe a range of concepts/fields and follows research trends

cept of data ecosystems is used to label familiar concepts in new guise. It seems such repackaging of established concepts as data ecosystems to give them an innovative character is misuse of the term. However, it is difficult to consistently challenge this characterization of data ecosystems, as the "softness" of the definition and the dual perspectives it originates from allow for these contemporary interpretations of the data ecosystem concept. Careful consideration is thus needed when denoting a use case or system as a data ecosystem. A critical question that must be asked is what is novel about a system that gives it an innovative character. This could be several of the concepts listed in Fig. 4. For example, a networked character seems essential to any data ecosystem. To denote a system as a data ecosystem, the innovative character should come from turning a platform into a network of platforms.

Important to denote here is that we were looking for literal descriptions like "Big Data Ecosystem" (Nadal et al. 2017), "Health Ecosystem" (Mahmood et al. 2023), "Data Space" (Möller et al. 2024) and so forth. Variations like "government Big Data Ecosystem" (for Big Data Ecosystem) (Shah et al. 2024) or "Cancer Ecosystem" (for Health Ecosystem") (Alfano et al. 2021) are also considered. However, a study that researched a "Big Data" data ecosystem use case but not explicitly named the concept a "Big Data Ecosystem" was not considered an explicit mention of the term and thus not included in the data displayed in Fig. 6. Again, not all studies describe data ecosystems in such a particular way, and some employ multiple such terms. Therefore, not all studies are represented, while others are represented more than once.

4.4 The diversity of actors and centrality of platforms in data ecosystems

Two parameters, actor diversity and platform centrality, provide further represent the use case mapping. On the one hand, we noticed data ecosystems that target either homogenous actors (e.g. only hospitals) or a diverse set of actors (e.g. hospitals, general practitioners, smartwatch manufacturers, and pharmacists). On the other hand, we observed ecosystems built around one central platform (e.g. portal of a national institute) or designed to allow platforms to be interchangeable (e.g. competing ride-sharing platforms). Figure 7 shows this interpretation of data ecosystem use cases. It must be noted this mapping is an interpretation of earlier results, rather than systematic mapping.

4.4.1 Homogenous actors & consists of multiple interchangeable platforms

In the first quadrant, no use cases were observed. It makes little sense to create data ecosystems for a homogenous group of actors that put effort into having interchangeable platforms. The effort would not outweigh the benefits, as this is a situation of platform competition, with limited to no cooperation opportunities (Rietveld and Schilling 2021). As no cases are observed, this shows scholars do not consider cases

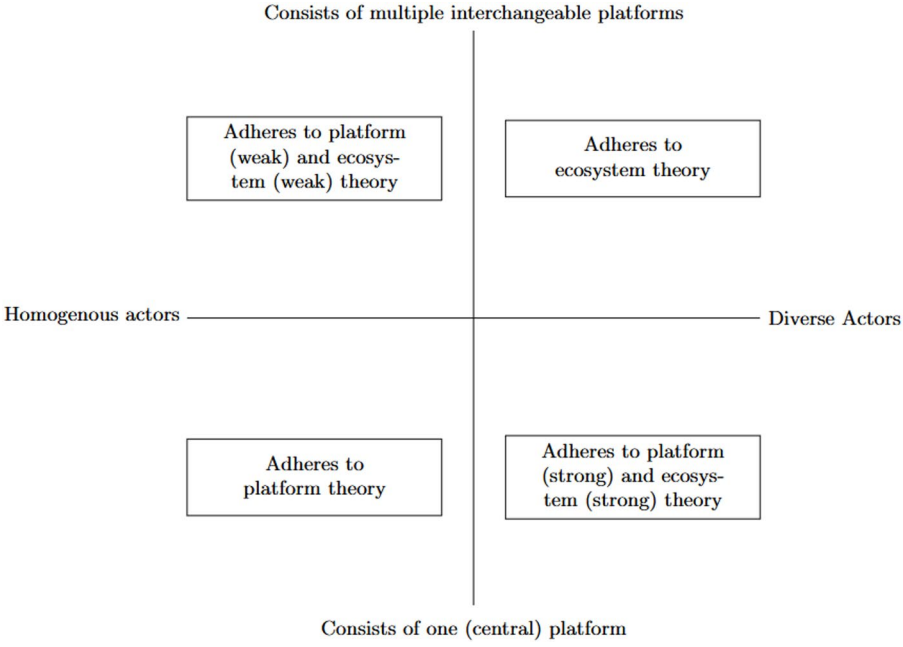


Fig. 7 Mapping the use cases into four quadrants (see Sect. 3.4), on the axes of actor diversity (horizontal) and platform centrality (vertical). Actor diversity concerns who is an actor in the ecosystem, either only homogenous actors (e.g. only hospitals) or a diverse set of actors (e.g. hospitals, general practitioners, pharmacists). Platform centrality deals with dependency on certain components. Either built around one central platform (e.g. portal of a national institute) or allow multiple platforms to participate (e.g. competing ride-sharing platforms)

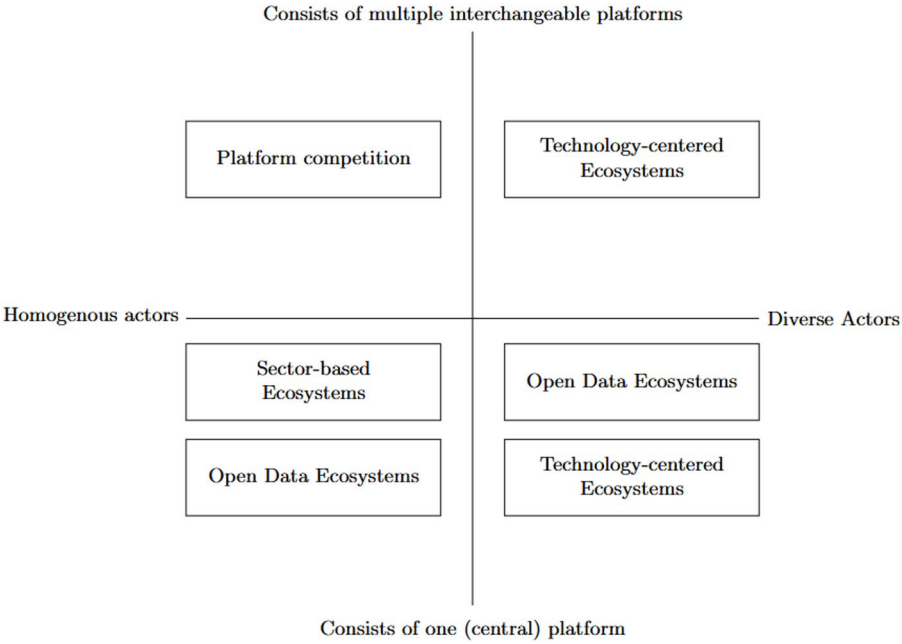


Fig. 8 The different quadrants of Fig. 7 on which the data ecosystem use cases found in literature are mapped. This mapping shows that what is considered a data ecosystem in practice (which use cases must represent) is not uniquely defined. Similar to the theoretical definitions discussed in this paper, the same problem applies to the use cases studied in data ecosystem literature. Coming back to the insights on the coding trees, this mapping also reveals that what is considered a platform in a data ecosystem context takes many forms. Again, adding to the problem that there is conceptual unclarity about what is, and is not a data ecosystem

with homogenous actors & multiple interchangeable platforms that serve the same purpose as data ecosystems.

4.4.2 Homogenous actors & consists of one (central) platform

Given that platforms are seen as technical implementations that are part of ecosystems, an argument can be made that when there is only one platform, the focus is on optimizing the platform's value gain and thus this quadrant adheres to platform theory. Two types of data ecosystem use cases found in the literature are mapped to this quadrant. The sector specific open data initiatives, and several of the sector-based data ecosystems. Lnenicka et al. (2022a), for example, identify smart city open data initiatives for which the open data platform is the center of the data infrastructure of the ecosystem. Another example are the open health data ecosystems (Heijlen and Crompvoets 2021; Shadbolt et al. 2022) which are open data portals where open health data is centrally accessible, that depend on government datasets and funding.

Concerning the homogenous actors, these authors observe that the current use is limited in diversity to research institutes (e.g. pharmaceutical companies). A portal is typically built, to solve problems that specific stakeholders currently face, rather than

focus on reaching new actors in the ecosystem. Therefore, not incentivizing a range of actors to form a broader ecosystem.

4.4.3 Diverse actors & consists of multiple interchangeable platforms

As Fig. 4 shows, ecosystem theory is about changing environments, in which diverse actors among which platforms cooperate and coexist. Accordingly, this quadrant has strong links to ecosystem theory. The use case clusters in this quadrant are those dealing with ODEs and technologies. They operate on a more fundamental level, where abstraction is made of use case specifics and the focus is on interoperable data sharing architecture technologies (Czvetkó and Abonyi 2023). Platforms are then seen as implementations of this architecture, and span multiple activities like clearinghouses, data brokers, identity providers and connectors. Another example is using blockchain to implement a distributed data sharing architecture (Xu et al. 2023). In line with this quadrants theme, the platform and sector agnostic nature is emphasized. Scheider et al. (2023a) introduce the concept of a meta-dimension of a service platform to personal data markets technology. This service platform represents the system architecture and technology choices. They consider multiple platform operators that provide different services for data sharing, showing the interchangeable nature of platforms in these data ecosystems.

4.4.4 Diverse actors & consists of one (central) platform

This quadrant complies with both the platform and ecosystem background of data ecosystems. Taking the central platform elements to adhere to platform theory, and the diverse actor elements from ecosystem theory, this quadrant describes elements from both backgrounds.

In this final quadrant there are several sector-based data ecosystems and Open Government Data platforms. These examples initiate the ecosystem, typically from a central platform, but aim their efforts at maximizing the data use for a diverse set of actors by means of diverting governance power in the ecosystem (Linåker and Runeson 2020). The platform is then seen as more of a central data-hub that must be accessible to all actors, rather than as a means to orchestrate the ecosystem (Aaen et al. 2022; Ramalli and Pernici 2023; Hawrylycz et al. 2023; Flores-Toro et al. 2023), These provide an example of how the notion of platforms is stretched, as for example Open Source projects are seen as possible platform providers. Essentially, going beyond the traditional view that platforms are centered around a vendor with central power.

5 Discussion

5.1 Main findings

Concerning our first research aim (1), we can conclude that literature has landed on a "most accepted" definition, namely the one conceptualized by Oliveira et al. (2019).

It is therefore advisable to use this common definition. However, the problem of the broadness of this definition remains, therefore, we call upon future studies to describe why a use case fits the definition rather than focusing on the exact definition.

The six key concepts provide answers to the second aim of this research (2). Figure 4 shows that data ecosystems can be approached through the lens of both platforms and ecosystems. While they are ingrained in the definition of data ecosystems, the weight and focus given to them varies on a case-by-case basis. For example, a standardization body publishing vocabularies for data exchange can emphasize the ecosystem perspective, as the use and success of the standard depends on a diverse ecosystem with many co-existing actors, including platforms (Woods et al. 2023). While a data-consumer platform, like a ride-sharing platform, can focus on the platform perspective, as interactions on the central platform create value for them (Lee et al. 2020). This emphasizes the direct engagement between platform users and broader ecosystem complementors. While a platform-centric approach focuses on the continued existence and dominance of the platform itself, its complementors are crucial for platform survival. An overly dominant platform can stifle the broader ecosystem, analogous to how a single species can overrun an ecosystem, reducing diversity and innovation potential.

Finally, the third research aim (3) was to clarify what is considered a data ecosystem instance. The broadness of data ecosystem definitions make it applicable across various sectors and contexts. Figure 6 shows that applicability also varies over time, following trends. Therefore, there is **no one-on-one applicability of definition to use case**. In particular, the positioning of platforms is remarkable. In theory platforms are considered part of data ecosystems, however all structure is lost when examining what is considered a platform in use cases. For example, a data portal (Linåker and Runeson 2020) or a common standard (Czvetkó and Abonyi 2023) are both seen as platforms in ecosystems, one refers to an actor, while the other is conceptual.

Concluding, there is little (call for) argumentation on **why** something is considered a data ecosystem. A likely reason is that there is no clear argument to be made why your system would not be a data ecosystem if it deals with sharing data among different participants, given the "softness" of its definitions. Therefore, we argue that **the data ecosystem term is in practice commonly (mis-)used to revitalize established concepts or systems as innovative**.

5.2 Theoretical implications

From a theoretical perspective, the most accepted definition of data ecosystems marks a major step forward, although the definition is still conceptually vague and diluted. This reflects vague boundaries between platform and ecosystem theory, on which the data ecosystem definitions are based. As Figs. 7 and 8 reveal, the concept of data ecosystems is all too often used to reformulate existing concepts, rather than to present a fundamentally novel idea. Thus, it is infeasible to alleviate the conceptual unclarity at the level of data ecosystems by means of a new definition. Rather, one should delineate the boundaries of what constitutes a data ecosystem by drawing on the six identified key concepts (Fig. 4) and showing how these are operationalized in their specific context. The study thus provides what is commonly understood

and accepted as a data ecosystem, how it is typically conceptualized and what this conceptualization means by examining how they are operationalized (e.g. how collaboration is supported). Both newcomers and knowledgeable readers can use it as a reference work on commonly accepted views on data ecosystems. However, a critical theoretical assessment of these accepted definitions and conceptualizations of data ecosystems is required.

There are multiple views on theory in information systems, with different purposes (Gregor 2006). Here we have assessed mainly the viewpoint of "explanation" theory as described by Gregor (2006), that has the goal to "say what is, how, why, when and where". In this work, this means how data ecosystems are described and conceptualized to understand and explain the recent evolutions in information systems for data sharing. From this perspective, we uncover what is accepted as a data ecosystem definition and use case, how it is conceptualized coming from different background theories, why there is a need for the data ecosystem concept, and finally when and where data ecosystems popped up by mapping and categorization of the use cases in literature. This was identified as where geographically or industry-wise these cases were observed and the periods in which certain trends arose in these cases. Having established a theoretical framework of what we investigated here as theory, the criteria proposed by Weber (2012) can be used to evaluate the commonly accepted data ecosystems conceptualization.

As Weber (2012) argues, theories should be evaluated on two levels: the parts of the theory (constructs, associations and states they cover) and the theory as a whole. Particularly important to our discussion is the emphasis Weber (2012) puts on how a theory's parts must be precisely described because they determine the phenomena it is intended to cover. This is exactly the issue with the lack of conceptual clarity of data ecosystems. More specifically, the constructs, which represent an attribute in general of some class of things, which make up data ecosystem conceptualization are not defined precisely. This issue results in the exact nature of the things the theory covers to be unclear (Weber 2012; Zhang et al. 2016), as we have observed mapping data ecosystem use cases. Another construct issue observed is that constructs can be proxies of one another, like the interconnection and interdependence, and the self-organizing constructs of data ecosystems. While there is a distinction made based on the economic (interconnection and interdependence) and social (self-organizing) background that the concepts are linked to, both refer to the same underlying property of actor behavior in relation to others in the ecosystem. As denoted by Weber (2012), such proxies result in tautological association. However, they also stress the difficulty of evaluating overlap in properties as we have experienced here. A similar view is found in the work of Zhang et al. (2016) stating logical relationships among constructs must be explained and preferably avoided but are inherent to exhaustive description of concepts.

On evaluating the theory as a whole, Weber (2012) mentions several criteria: importance, novelty, parsimony, level, and falsifiability. First, the importance of the theory seems obvious given the (still growing) amount of literature on data ecosystems was found. Second, for a theory to be novel, it must shift paradigms. Regarding data ecosystems we dispute the novelty given the link to existing platform and ecosystem theory, both in conceptualization, use cases, and its susceptibility to trends.

Thirdly, the parsimony or efficiency (Zhang et al. 2016) of data ecosystem theory, expressed in achieving good levels of explanatory power using a small number of constructs, is an issue for data ecosystems. By aggregating conceptualization from all the studies, we still need several concepts from platform, ecosystem and data ecosystem theory, to grasp data ecosystems. Fourthly, regarding the level, data ecosystems are described on a macro-level, which suffers from imprecise construct definitions (Weber 2012). This shows once more specific phenomena are to be described in the use cases as the general descriptions of data ecosystems offer only limited explanatory power. Finally, we do believe the data ecosystem theory is susceptible to falsification because empirical outcomes can possibly be finessed using the theory. This is due to a lack of clear test conditions, as both the conceptualization and use case analysis shows. It remains hard to argue why a system, like a platform implementation, should not be considered a data ecosystem. However, we do not believe the theory itself is falsified, or false "tests" for data ecosystems are common but rather think this is a result of the imprecise constructs. Mainly, scholars looking into contributing to data ecosystem theory must be aware of the precarity of testing the conceptual clarity of data ecosystems.

5.3 Practical & policy implications

The findings in literature on the conceptualization of data ecosystems in relation to platform and ecosystem theory highlight the need for conscious operationalization of these different key concepts when setting up, or operating, a data ecosystem. Having a clear delineation of these concepts allows for better identification of threats and opportunities to the governance and business models of data ecosystems. The business models describe the different actors and how they offer their services in data ecosystem use cases (de Mildt et al. 2025). The governance models describe the processes that regulate data access, control and accountability in the ecosystem among actors (Lis and Otto 2020). For example, a sustainable business model, from a platform perspective, might miss a crucial element in their governance model to allow the platform to coexist within a data ecosystem. Reconciliation is needed if a platform emphasizes data governance and architectural control mechanisms, whereas its broader ecosystem focusses primarily on relational control mechanisms (e.g. shared values).

The issues on applicability to use cases present another business model challenge where the value of use cases is possibly diminished by labeling them as a data ecosystem (de Mildt et al. 2025). What we mean by this is that the data ecosystem label will result in possibly overclaiming value or taking focus away from tangible value a use case offers. Packaging a platform as a data ecosystem, which does not offer anything innovative, like solving simple data access or aggregation inefficiency, takes away the focus of the legitimate value that a way to access data can offer (e.g. use cases for Open Data platforms). Basically, overselling a system as a data ecosystem leads to innovative goals not to be reached, and investment to be wasted resulting in duped investors and damaged reputation and trust in both the concept and its implementors. On the other hand, overengineering a system to fit the data ecosystem term is also observed. Extensively incorporating data sharing standards and intermediaries to

solve one-on-one data exchanges complicates the solution, making use cases almost impossible to resolve. Introducing intermediaries, which possibly opens new opportunities to share data, also introduces new liability. Data provenance becomes more complicated, as yet another party handles data, inducing complexity to adhere to data protection law (Fierens et al. 2025).

To further strengthen the dimensions of funding being misallocated, as mentioned in the introduction, several relevant policies fit into the strategy the EU has for creating a single data market (and a similar argument can be made for non-EU strategies and initiatives). From this perspective, funding is provided to initiatives and organizations that support this strategy. Not surprisingly, the term data ecosystem is used by these initiatives to do exactly that: label systems as data ecosystems to fit the development of this strategy. It is apparent that if a non-innovative initiative receives funding by packaging existing platforms or ecosystems as data ecosystems, taxpayers' money is wasted. In practice a definition, or at least an assessment tool, of data ecosystems that (while seemingly an impossible feat) encompasses this innovation aspect is needed. Such assessment must stress having a truly innovative character regarding elements like cooperations in the ecosystem (e.g. parties now cooperate that were not able to before). The data exchanges in the ecosystem should not just be one-way traffic, and exchanging parties must be aware of a two-way data exchange. Part of the data ecosystem literature would argue platforms solely basing their business model on gathering, for example, behavioral web analytics data on end users are not meant to be labeled as a data ecosystem due to the lack of an innovative character. The issue that remains, however, is that what specifically is innovative is inherently open to interpretation. From analyzing literature and extracting practical use cases on data ecosystems, we learned that the concept of a data ecosystem represents an innovative idea, more than a type of system. Therefore, critical application and evaluation of the term would be a significant advancement to current practice.

From a policy makers perspective, understanding common conceptualizations of data ecosystems allows analyzing how gatekeeping and monopolistic risks vary if different weight is attributed to platform and ecosystem dynamics. The former entails forms of dominance through self-reinforcing network effects, while the latter allows for the co-existence of competition, collaboration and co-innovation. This is of relevance to legislation such as the Digital Markets Act (Parliament and Commission of the European Union 2022). For example, understanding whether a central platform is controlling the data ecosystem, allows for retribution to the platform as a possibly successful measure. Additionally, an organization might act as a data provider in one scenario and as a data consumer in another. Such circular use of data necessitates a flexible governance structure which rigid policies may fail to capture. Literature suggests that this may be an outcome of the DGA, where not having to comply with the DGA provides an advantage, rather than a restriction on operations (Verstraete et al. 2023), showing the relevance of data ecosystem conceptualization for policy makers.

6 Conclusion & future research

This paper aimed to (1) investigate how data ecosystems are conceptualized and whether there is a common definition, (2) identify the key concepts from which these definitions draw and how they co-shape the concept of data ecosystems, and (3) evaluate whether these definitions can be applied to concrete use cases in a meaningful way. An answer to these questions was provided as follows. (1) There is a generally accepted definition of data ecosystems, as provided by Oliveira et al. (2019). However, the need for clearly defining what is considered a data ecosystem is still valid, given a broad range of systems fit this definition. (2) Six key concepts are central to, and co-shape, the definition of data ecosystems: interconnection and interdependence, value networks, ecosystem metaphor, sociotechnical, actor networks and self-organization. These six concepts find their origin in both platform and ecosystem theory, providing relevant context for data ecosystem analysis and operationalization. (3) These definitions are applied differently across various contexts, which can be grouped into three distinct categories: Open Data Ecosystems, Sector-based Ecosystems, and Technology-centered Ecosystems. This mapping revealed that the data ecosystem concept is susceptible to trends and often used to repackage systems as innovative without adequate justification. Concluding, this study established that scoping data ecosystems is more about defining the right context and perspective for each system, and making sure systems can co-exist, than it is about setting up an exhaustive list of boundaries or exclusion criteria.

Regarding future research, these platform and ecosystem perspectives need to be linked to other aspects of data ecosystems. Having an influence on governance and business models behind data ecosystems requires these perspectives to be considered when setting them up for success. One approach can be to identify contingency factors, which are the specific conditions or variables that affect how data ecosystem governance should be structured (Lee et al. 2018; Costabile 2024). The provided lens, which approaches data ecosystems from both a platform and ecosystem perspective, provides a basis for further research on identifying, understanding, and addressing what contingency factors are vital for designing effective data ecosystem governance models that can adapt to contextual complexities and challenges.

Acknowledgements Obfuscated for review.

Author contributions M.d.M. and S.V.D. wrote the main manuscript text and performed the survey selection and coding. S.V. and D.C. reviewed and provided insights and research ideas during the selection, coding and the building of the discussion on the findings. All authors reviewed the manuscript. M.d.M. and S.V. D. performed the revision of the manuscript.

Data availability The data on the selection of the studies and the coding is available on https://github.com/TechnoEconomics/[paper identifier] (statement to be updated upon acceptance).

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Aaen J, Nielsen JA, Carugati A (2022) The dark side of data ecosystems: A longitudinal study of the DAMD project. Eur J Inform Syst 31(3):288–312. https://doi.org/10.1080/0960085X.2021.1947753

Acs ZJ, Song AK, Szerb L, Audretsch DB, Komlósi É (2021) The evolution of the global digital platform economy: 1971–2021. Small Bus Econ 57(4):1629–1659. https://doi.org/10.1007/s11187-021-00561-x

Aguilera U, Peña O, Belmonte O, López-de Ipiña D (2017) Citizen-centric data services for smarter cities. Future Generation Comput Syst. https://doi.org/10.1016/j.future.2016.10.031

Aho B (2024) Data communism: Constructing a national data ecosystem. Big Data Soc. https://doi.org/10.1177/20539517241275888

Alfano C, Suarez-Almazor M, Rodriguez M, Palos G, Gilmore K (2021) Innovating the personalization of stratified survivorship care pathways: Using a cancer data ecosystem to improve care access, outcomes, efficiency, and costs. J Natl Cancer Inst Monogr. https://doi.org/10.1093/jncimonographs/lgaa011

Allemang D (2019) Sustainability in data and food. Data Intell. https://doi.org/10.1162/dint_a_00005

Altendeitering M, Guggenberger TM, Möller F (2024) A design theory for data quality tools in data ecosystems: Findings from three industry cases. Data Knowl Eng. https://doi.org/10.1016/j.datak.2024.102333

Ammann J, Hess T (2025) To sell, to donate, or to barter? Value creation and capture through business model types in decentralized data ecosystems. Electron Markets 35(36). https://doi.org/10.1007/s12525-025-00775-x

Antolini F, Terraglia I, Cesarini S (2024) Integrating multiple data sources to measure sustainable tourism in Italian regions. Socio-Economic Plann Sci. https://doi.org/10.1016/j.seps.2024.101959

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

Baldacci E, Braaksma B, Gálvez A, Giannakouris K, Olmos BG, Rivière P, Scannapieco M, Vermeer A, Vertanen V (2022) Innovation during the COVID-19 crisis: Why it was more critical for official statistics than ever. Stat J IAOS 38(2):399–412. https://doi.org/10.3233/SJI220937

Bao A-R-H, Liu Y, Dong J, Chen Z-P, Chen Z-J, Wu C (2022) Evolutionary game analysis of co-opetition strategy in energy big data ecosystem under government intervention. Energies 15(6):2066. https://doi.org/10.3390/en15062066

Barns S (2018) Smart cities and urban data platforms: Designing interfaces for smart governance. City Cult Soc 12:5–12. https://doi.org/10.1016/j.ccs.2017.09.006

Begany GM, Martin E (2017) An open health data engagement ecosystem model: Are facilitators the key to open data success? Proc Association Inform Sci Technol. https://doi.org/10.1002/pra2.2017.14505401091

Blaiszik B, Ward L, Schwarting M, Gaff J, Chard R, Pike D, Chard K, Foster I (2019) A data ecosystem to support machine learning in materials science. MRS Commun. https://doi.org/10.1557/mrc.2019.118

Booth P, Navarrete T, Ogundipe A (2021) Museum open data ecosystems: A comparative study. J Doc 78(4):761–779. https://doi.org/10.1108/JD-05-2021-0102

Boulet V (2022) How to build an identifiers policy: The BnF use case. JLIS it. https://doi.org/10.4403/jlis.it-12768

Boyles RR, Thessen AE, Waldrop A, Haendel MA (2019) Ontology-based data integration for advancing toxicological knowledge. Curr Opin Toxicol. https://doi.org/10.1016/j.cotox.2019.05.005

Briscoe G, De Wilde P (2006) Digital ecosystems: Evolving service-orientated architectures. In Proceedings of the 1st International Conference on Bio Inspired Models of Network, Information and Computing Systems. https://doi.org/10.4108/bionetics.2006.184
Calzada I, Almirall E (2020) Data ecosystems for protecting European citizens' digital rights. Transforming Government: People Process Policy 14(2):133–147. https://doi.org/10.1108/TG-03-2020-0047
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. https://doi.org/10.1017/dap.2023.18
Casanovas P, Mendelson D, Poblet M (2017) A linked democracy approach for regulating public health data. Health Technol. https://doi.org/10.1007/s12553-017-0191-5
Chamorro-Padial J, García R, Gil R (2024) A systematic review of open data in agriculture. Comput Electron Agric. https://doi.org/10.1016/j.compag.2024.108775
Chen C, Howe J, Kariotis T, Jackson S (2024) Open data, workplace relations law compliance, and digital regulation. Data Policy. https://doi.org/10.1017/dap.2024.18
Chevallier T, Lauer J, Renso C, Blanco-Justicia A, De Ryck D, Papacharalampous AE (2023) Making it easy for transport stakeholders to share mobility data. Transp Res Procedia. https://doi.org/10.1016/j.trpro.2023.11.711
Chu I, Miller R, Mathews I, Vala A, Sept L, O'Hara R, Rehkopf DH (2024) Fair enough: Building an academic data ecosystem to make real-world data available for translational research. J Clin Translational Sci. https://doi.org/10.1017/cts.2024.530
Common European Data Spaces (2024) European Commission. https://digital-strategy.ec.europa.eu/en/policies/dataspaces
Corbett J, Templier M, Townsend H, Takeda H (2020) Integrating across sustainability, political, and administrative spheres: A longitudinal study of actors' engagement in open data ecosystems in three Canadian cities. Commun Association Inform Syst 47:28. https://doi.org/10.17705/1CAIS.04728
Côrte-Real N, Ruivo P, Oliveira T (2020) Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? Inform Manager. https://doi.org/10.1016/j.im.2019.01.003
Costabile C (2024) Digital platform ecosystem governance of private companies: Building blocks and a research agenda based on a multidisciplinary, systematic literature review. Data Inform Manage 8(1):100053. https://doi.org/10.1016/j.dim.2023.100053
Cruz-Almeida Y, Mehta B, Haelterman NA, Johnson AJ, Heiting C, Ernberg M, Orange DE, Lotz M, Boccanfuso J, Smith SB, Pela M, Boline J, Otero M, Allen KD, Pérez D, Donnelly CR, Almarza AJ, Olmer M, Balkhi H, Wagenaar J, Martone ME (2024) Clinical and biobehavioral phenotypic assessments and data harmonization for the RE-JOIN research consortium: Recommendations for common data element selection. Neurobiol Pain. https://doi.org/10.1016/j.ynpai.2024.100163
Curry E, Sheth A (2018) Next-generation smart environments: From system of systems to data ecosystems. IEEE Intell Syst 33(3):69–76. https://doi.org/10.1109/MIS.2018.033001418
Curry E, Derguech W, Hasan S, Kouroupetroglou C, ul Hassan U (2019) A real-time linked dataspace for the internet of things: enabling pay-as-you-go data management in smart environments. Future Generation Comput Syst 90:405–422
Czvetkó T, Abonyi J (2023) Data sharing in Industry 4.0 – AutomationML, B2MML and International Data Spaces-based solutions. J Industrial Inform Integr 33:100438. https://doi.org/10.1016/j.jii.2023.100438
D'Hauwers R, Vandercruysse L (2025) Competitive advantage and personal data ecosystems: A typology of personal data control constellations. J Theoretical Appl Electron Commer Res 20(1):8. https://doi.org/10.3390/jtaer20010008
Dawes SS, Vidiasova L, Parkhimovich O (2016) Planning and designing open government data programs: An ecosystem approach. Government Inform Q 33(1):15–27. https://doi.org/10.1016/j.giq.2016.01.003
de Mildt M, Verbrugge S, Colle D (2025) A market analysis on data ecosystem initiators and their value propositions in different ecosystems. Telecomm Policy 49(3):102910. https://doi.org/10.1016/j.telpol.2024.102910
De Reuver M, Sørensen C, Basole RC (2018) The digital platform: A research agenda. J Inform Technol 33(2):124–135. https://doi.org/10.1057/s41265-016-0033-3
Denecke K, Romero OR, Merolli M, Miron-Shatz T, Gabarron E, Petersen C (2023) How participatory health informatics catalyzes one digital health. Yearb Med Inform. https://doi.org/10.1055/s-0043-1768727

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. https://doi.org/10.3390/en13020444

Donida B, da Costa CA, Scherer JN (2021) Making the COVID-19 pandemic a driver for digital health: Brazilian strategies. JMIR Public Health Surveillance. https://doi.org/10.2196/28643

DSSC (2024) Introduction – Key Concepts of Data Spaces. https://dssc.eu/space/bv15e/766061351/Introduction+-+Key+Concepts+of+Data+Spaces

Eastwood C, Edwards J, Turner J (2021) Anticipating alternative trajectories for responsible agriculture 4.0 innovation in livestock systems. Animal. https://doi.org/10.1016/j.animal.2021.100296

Ertürkmen GBL, Sınacı AA, Namlı T, Boonstra M, Lekadir K, Gkontra P, Chronaki C, Facile R, Baker R, Kush R (2024) An assessment of the European patient summary for clinical research: A case study in cardiology. Front Med. https://doi.org/10.3389/fmed.2024.1481551

European Union (2025) Open Data Maturity 2024. data.europa.eu. https://data.europa.eu/en/publications/open-data-maturity

European Commission (2024) European strategy for data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

Exner TE, Papadiamantis AG, Melagraki G, Amos JD, Bossa N, Gakis GP, Charitidis CA, Cornelis G, Costa AL, Doganis P, Farcal L, Friedrichs S, Furxhi I, Klaessig FC, Lobaskin V, Maier D, Rumble J, Sarimveis H, Suarez-Merino B, Vázquez S, Wiesner MR, Afantitis A, Lynch I (2023) Metadata stewardship in nanosafety research: Learning from the past, preparing for an on-the-fly FAIR future. Front Phys 11. https://doi.org/10.3389/fphy.2023.1233879

Fang J, Zhao L (2024) Synergistic evolution mechanisms for improving open government data ecosystems using the Haken model. Libr Inform Sci Res. https://doi.org/10.1016/j.lisr.2024.101284

Fang J, Zhao L, Li S (2024) Exploring open government data ecosystems across data, information, and business. Government Inform Q. https://doi.org/10.1016/j.giq.2024.101934

Farrell E et al (2023) European data spaces-scientific insights into data sharing and utilisation at scale. Technical Report JRC129900, Joint Research Centre (Seville site)

Fassnacht M, Leimstoll J, Benz C et al (2024) Data sharing practices: The interplay of data, organizational structures, and network dynamics. Electron Markets 34(47). https://doi.org/10.1007/s12525-024-00732-0

Fenwick M, Jurčys P (2023) Building a 'green data' future: How a human-centric approach to data and nudges can help fight climate change. SSRN Electron J. https://doi.org/10.2139/ssrn.4326718

Ferretti A, Vayena E (2022) In the shadow of privacy: Overlooked ethical concerns in COVID-19 digital epidemiology. Epidemics. https://doi.org/10.1016/j.epidem.2022.100652

Feyzbakhsh OA-B, Babalhavaeji F, Nezafati N, Hariri N, Nooshinfard F (2022) Presenting an open data management (ecosystem) model with developing innovative information flow approach in Iranian knowledge-based companies. Aslib J Inform Manage 74(3):458–494. https://doi.org/10.1108/AJIM-07-2021-0186

Fierens M, Pandit HJ, Tamo-Larrieux A, Garcia K (2025) A Solid use case to empower and protect data subjects: Responsibilities under GDPR for governance of personal data stores. Comput Law Secur Rev 57:106133. https://doi.org/10.1016/j.clsr.2024.106133

Flores-Toro JA, Jagu S, Armstrong GT, Arons DF, Aune GJ, Chanock SJ, Hawkins DS, Heath A, Helman LJ, Janeway KA, Levine JE, Miller E, Penberthy L, Roberts CWM, Shalley ER, Shern JF, Smith MA, Staudt LM, Volchenboum SL, Zhang J, Zenklusen JC, Lowy DR, Sharpless NE, Auvil G, Kerlavage JM, Widemann AR, Reaman BC, Kibbe GH, Doroshow WA, on behalf of Childhood Cancer Data Initiative Working Groups (2023) The Childhood Cancer Data Initiative: Using the power of data to learn from and improve outcomes for every child and young adult with pediatric cancer. J Clin Oncol 41(24):4045–4053. https://doi.org/10.1200/JCO.22.02208

Fraga KJ, Huang YJ, Ramelot TA, Kendary ALA, Li E, Korf I, Montelione GT (2022) SpecDB: A relational database for archiving biomolecular NMR spectral data. J Magn Reson. https://doi.org/10.1016/j.jmr.2022.107268

Francia M, Gallinucci E, Golfarelli M, Leoni AG, Rizzi S, Santolini N (2021) Making data platforms smarter with MOSES. Future Generation Comput Syst. https://doi.org/10.1016/j.future.2021.06.031

Frigeri M, King L, Musidlowska M, Peeters B (2024) *From Data Silos to Data Ecosystems: The role of Data Spaces Support Centre in addressing the legal challenges of the future single market for data* [Blog post]. CiTiP, KU Leuven. https://www.law.kuleuven.be/citip/blog/from-data-silos-to-data-ecosystems-the-role-of-data-spaces-support-centre-in-addressing-the-legal-challenges-of-the-future-single-market-for-data/

Fu Z, Yao S, Tiwari AK, Mohiuddin M, Zhong K, Haseeb M, Liu Y (2024) The impact of industrial agglomeration on the synergistic evolution of the energy big data ecosystem: Empirical findings from China. J Environ Planning Manage. https://doi.org/10.1080/09640568.2024.2413879

Geisler S, Vidal M-E, Cappiello C, Lóscio BF, Gal A, Jarke M, Lenzerini M, Missier P, Otto B, Paja E, Pernici B, Rehof J (2021) Knowledge-driven data ecosystems toward data transparency. J Data Inform Qual 14(1):3:1–312. https://doi.org/10.1145/3467022

Geldof T, Huys I, Van Dyck W (2019) Real-world evidence gathering in oncology: The need for a biomedical big data insight-providing federated network. Front Med 6:43. https://doi.org/10.3389/fmed.2019.00043

Gelhaar J, Groß T, Otto B (2021) A taxonomy for data ecosystems. In Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS). https://doi.org/10.24251/HICSS.2021.100

Ghazinoory S, Nasri S, Dastranj R, Sarkissian A (2021) Bio to bits: The Millennium Ecosystem Assessment (MA) as a metaphor for big data ecosystem assessment. Inform Technol People 35(2):835–858. https://doi.org/10.1108/ITP-08-2020-0579

Given LM, Polkinghorne S, Cattlin J (2023) Structural elements and spheres of expertise: Creating a healthy ecosystem for cultural data initiatives. J Association Inform Sci Technol. https://doi.org/10.1002/asi.24849

Gordon-Tapiero A, Wood A, Ligett K (2022) The case for establishing a collective perspective to address the harms of platform personalization. In Proceedings of the 2022 Symposium on Computer Science and Law (CSLAW '22) (pp. 119–130). Association for Computing Machinery. https://doi.org/10.1145/3511265.3550450

Gout AM, Arunachalam S, Finkelstein DB, Zhang J (2021) Data-driven approaches to advance research and clinical care for pediatric cancer. Biochim et Biophys Acta (BBA) - Reviews Cancer. https://doi.org/10.1016/j.bbcan.2021.188571

Grassi L, Lanfranchi D (2022) RegTech in public and private sectors: The nexus between data, technology and regulation. J Industrial Bus Econ 49(3):441–479. https://doi.org/10.1007/s40812-022-00226-0

Gregor S (2006) The nature of theory in information systems. MIS Q 30(3):611–642. https://doi.org/10.2307/25148742

Gröger C (2021) There is no AI without data. Commun ACM 64(11):98–108. https://doi.org/10.1145/3448247

Grossman R (2018) Progress toward cancer data ecosystems. Cancer J 24(1):1–5. https://doi.org/10.1097/PPO.0000000000000318

Guggenberger TM, Möller F, Haarhaus T, Gür I, Otto B (2020), June Ecosystem types in information systems. In Proceedings of the European Conference on Information Systems (ECIS)

Guggenberger TM, Langdon SC, Otto B (2025) Data spaces as meta-organisations. Eur J Inform Syst 1–21. https://doi.org/10.1080/0960085X.2025.2451250

Gupta A, Panagiotopoulos P, Bowen F (2020) An orchestration approach to smart city data ecosystems. Technol Forecast Soc Chang 153:119929. https://doi.org/10.1016/j.techfore.2020.119929

Hamed N, Rana O, Orozco-ter Wengel P, Goossens B, Perera C (2024) A comparison of open data observatories. J Data Inform Qual. https://doi.org/10.1145/3705863

Hansen CJ, Cespedes LMS, Oyola LT, Dimakos XK, Walsh B, Bueno RS, Kabo-Bah AT, Seidu O, Nielsen VO (2024) Collaboration between national statistical offices and academia: Benefits, conditions, areas of collaboration and practical level experience in countries. Stat J IAOS. https://doi.org/10.3233/SJI230117

Hawrylycz M, Martone ME, Ascoli GA, Bjaalie JG, Dong H-W, Ghosh SS, Gillis J, Hertzano R, Haynor DR, Hof PR, Kim Y, Lein E, Liu Y, Miller JA, Mitra PP, Mukamel E, Ng L, Osumi-Sutherland D, Peng H, Ray PL, Zingg B (2023) A guide to the BRAIN Initiative Cell Census Network data ecosystem. PLoS Biol 21(6):e3002133. https://doi.org/10.1371/journal.pbio.3002133

Hayashi T, Ishimura G, Ohsawa Y (2021) Structural characteristics of stakeholder relationships and value chain network in data exchange ecosystem. IEEE Access 9:52266–52276. https://doi.org/10.1109/ACCESS.2021.3070017

Heijlen R, Crompvoets J (2021) Open health data: Mapping the ecosystem. Digit Health 7:20552076211050167. https://doi.org/10.1177/20552076211050167

Hein A, Schreieck M, Riasanow T, Setzke DS, Wiesche M, Böhm M, Krcmar H (2020) Digital platform ecosystems. Electron Markets 30(1):87–98. https://doi.org/10.1007/s12525-019-00377-4

Heltweg P, Riehle D (2023) A systematic analysis of problems in open collaborative data engineering. ACM Trans Social Comput. https://doi.org/10.1145/3629040

Hidalgo N, Rosas E, Vasquez C, Wladdimiro D (2018) Measuring stream processing systems adaptability under dynamic workloads. Future Generation Comput Syst 88:1–14. https://doi.org/10.1016/j.future.2018.05.084

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-Information 7(12):456. https://doi.org/10.3390/ijgi7120456

Howard C, Hyland-Wood B (2024) The promise and performance of data ecosystems: Australia's COVID-19 response. Policy Politics 52(1):108–130. https://doi.org/10.1332/03055736Y2023D000000001

Hrustek L, Tomicic Furjan M, Varga F, Dzidic A, Van Loenen B, Salamon D (2022) Towards digital innovation: Stakeholder interactions in agricultural data ecosystem in Croatia. Interdisciplinary Description Complex Syst 19(4):491–499. https://doi.org/10.7906/indecs.20.2.10

Hussein R, Griffin AC, Pichon A, Oldenburg J (2023) A guiding framework for creating a comprehensive strategy for mHealth data sharing, privacy, and governance in low- and middle-income countries (LMICs). J Am Med Inform Assoc 30(4):787–794. https://doi.org/10.1093/jamia/ocac198

Hutterer A, Krumay B, Mühlburger M (2023) What constitutes a dataspace? Conceptual clarity beyond technical aspects

Hyysalo J, Dasanayake S, Hannu J, Schuss C, Rajanen M, Leppänen T, Doermann D, Sauvola J (2022) Smart mask – Wearable IoT solution for improved protection and personal health. Internet Things 18:100511. https://doi.org/10.1016/j.iot.2022.100511

Iacob N, Simonelli F (2020) Towards a European health data ecosystem. Eur J Risk Regul. https://doi.org/10.1017/err.2020.88

Immonen A, Kiljander J, Aro M (2020) Consumer viewpoint on a new kind of energy market. Electr Power Syst Res. https://doi.org/10.1016/j.epsr.2019.106153

International Data Spaces Association (2024) Retrieved June 2024, from https://internationaldataspaces.org/

Janev V, Vidal ME, Endris K, Pujic D (2021), June Managing knowledge in energy data spaces. In Companion Proceedings of the Web Conference 2021 (pp. 7–15). Association for Computing Machinery. https://doi.org/10.1145/3442442.3453541

Janev V, Vidal M-E, Pujíc D, Popadíc D, Iglesias E, Sakor A, Campa A (2022) Responsible knowledge management in energy data ecosystems. Energies 15(11):3973. https://doi.org/10.3390/en15113973

Jean-Quartier C, Rey Mazón M, Lovríc M, Stryeck S (2022) Collaborative data use between private and public stakeholders—A regional case study. Data 7(2):20. https://doi.org/10.3390/data7020020

Jussen I, Schweihoff J, Dahms V, Möller F, Otto B (2023) Data sharing fundamentals: Characteristics and definition. In Proceedings of the 56th Hawaii International Conference on System Sciences.

Jussen I, Möller F, Schweihoff J, Gieß A, Giussani G, Otto B (2024) Issues in inter-organizational data sharing: Findings from practice and research challenges. Data Knowl Eng 150:102280. https://doi.org/10.1016/j.datak.2024.102280

Kamariotou M, Kitsios F (2022) Bringing digital innovation strategies and entrepreneurship: The business model canvas in open data ecosystem and startups. Future Internet 14(5):127. https://doi.org/10.3390/fi14050127

Kapoor M, Ventura ES, Walsh A, Sokolov A, George N, Kumari S, Provart N, Cole B, Libault M, Tickle T, Warren W, Koltes J, Papatheodorou I, Ware D, Harrison PW, Elsik CG, Yordanova G, Burdett T, Tuggle CK (2024) Building a FAIR data ecosystem for incorporating single-cell transcriptomics data into agricultural genome to phenome research. Front Genet. https://doi.org/10.3389/fgene.2024.1460351

Karanja EM, Masupe S, Gasennelwe-Jeffrey M (2018) Challenge paper: Towards open datasets for internet of things malware. J Data Inform Qual 10(2). https://doi.org/10.1145/3230669

Kari A, Schurig T, Gersch M (2024) The emergence of a new European data economy: A systematic research agenda for health data spaces. SMR—Journal Service Manage Res 7(4):176–198

Keim-Malpass J, Kausch S (2023) Data science and precision oncology nursing: Creating an analytic ecosystem to support personalized supportive care across the trajectory of illness. Semin Oncol Nurs. https://doi.org/10.1016/j.soncn.2023.151432

Kemmer I, Keppler A, Serrano-Solano B, Rybina A, Ozdemir B, Bischof J, El Ghadraoui A, Eriksson JE, Mathur A (2023) Building a FAIR image data ecosystem for microscopy communities. Histochem Cell Biol 160(3):199–209. https://doi.org/10.1007/s00418-023-02203-7

Kenney M, Zysman J (2016) The rise of the platform economy. Issues Sci Technol 32(3):61

Khezr S, Yassine A, Benlamri R, Hossain MS (2022) An edge intelligent blockchain-based reputation system for IIoT data ecosystem. IEEE Trans Industr Inf 18(11):8346–8355. https://doi.org/10.1109/TII.2022.3174065

Khurshid M, Zakaria NH, Rashid A, Kazmi R, Shafique MN, Ahmad MN (2019) Analyzing diffusion patterns of big open data as policy innovation in public sector. Comput Electr Eng. https://doi.org/10.1016/j.compeleceng.2019.07.010
Kiesel R, Weiss A, Schmitt R (2021) Filling the semantics gap in industrial communication: Middleware + for an internet of production. IFAC-PapersOnLine. https://doi.org/10.1016/j.ifacol.2021.08.049
Kitsios F, Kamariotou M, Grigoroudis E (2021) Digital entrepreneurship services evolution: Analysis of quadruple and quintuple helix innovation models for open data ecosystems. Sustainability 13(21):12183. https://doi.org/10.3390/su132112183
Knieps G (2019) Internet of things, big data and the economics of networked vehicles. Telecomm Policy. https://doi.org/10.1016/j.telpol.2018.09.002
Konstantinidis I, Kapantai E, Michailidis A, Deligiannis A, Berberidis C, Magnisalis I, Peristeras V (2024) From document-centric to data-centric public service provision. Digit Government. https://doi.org/10.1145/3676279
Koren O, Hallin CA, Perel N, Bendet D (2019) Decision-making enhancement in a big data environment: Application of the k-means algorithm to mixed data. J Artif Intell Soft Comput Res. https://doi.org/10.2478/jaiscr-2019-0010
Kruk SR, Bush SR, Phillips MJ (2024) Federating 'aquaculture 4.0' for data-driven social and environmental sustainability. Mar Policy. https://doi.org/10.1016/j.marpol.2024.106355
Kumar V, Kumar K, Hussain M (2023) Enhanced generic framework for privacy preservation in internet of things networks. Procedia Comput Sci. https://doi.org/10.1016/j.procs.2023.12.103
Kyoda K, Itoga H, Yamagata Y, Fujisawa E, Wang F, Miranda-Miranda M, Yamamoto H, Nakano Y, Tohsato Y, Onami S (2024) SSBD: An ecosystem for enhanced sharing and reuse of bioimaging data. Nucleic Acids Res. https://doi.org/10.1093/nar/gkae860
Lee M-H (2019) Challenges of designing with open data: The case of cultural data. Archives Des Res. https://doi.org/10.15187/adr.2019.08.32.3.57
Lee SU, Zhu L, Jeffery R (2018) A contingency-based approach to data governance design for platform ecosystems
Lee A, Mackenzie A, Smith GJD, Box P (2020) Mapping platform urbanism: Charting the nuance of the platform pivot. Urban Plann 5(1):116–128. https://doi.org/10.17645/up.v5i1.2545
Legenvre H, Hameri A-P (2023) The emergence of data sharing along complex supply chains. Int J Oper Prod Manage 44(1):292–297. https://doi.org/10.1108/IJOPM-11-2022-0729
Li W, Raman L, Debnath A, Ahn M, Lin S, Krajewski AM, Shang S, Priya S, Reinhart WF, Liu Z, Beese AM (2024) Design and validation of refractory alloys using machine learning, CALPHAD, and experiments. Int J Refract Met Hard Mater. https://doi.org/10.1016/j.ijrmhm.2024.106673
Liang W, Liu F, Wang L, Zheng B, Sun Y (2023) Internet of cultural things: Current research, challenges and opportunities. https://doi.org/10.32604/cmc.2023.029641
Linåker J, Runeson P (2020) Collaboration in open government data ecosystems: Open cross-sector sharing and co-development of data and software. In G. Viale Pereira, M. Janssen, H. Lee, I. Lindgren, M. P. Rodríguez Bolívar, H. J. Scholl, & A. Zuiderwijk (Eds.), *Electronic Government* (pp. 290–303). Springer. https://doi.org/10.1007/978-3-030-57599-1_22
Linåker J, Runeson P (2021) How to enable collaboration in open government data ecosystems: A public platform provider's perspective. JeDEM: eJournal eDemocracy Open Government. https://doi.org/10.29379/jedem.v13i1.634
Linåker J, Runeson P (2022) Sustaining open data as a digital common – Design principles for common pool resources applied to open data ecosystems. In *Proceedings of the 18th International Symposium on Open Collaboration, OpenSym '22* (pp. 1–11). ACM. https://doi.org/10.1145/3555051.3555066
Lis D, Otto B (2020) Data governance in data ecosystems – Insights from organizations. AMCIS 2020 Proceedings, 12. https://aisel.aisnet.org/amcis2020/strategic_uses_it/strategic_uses_it/12
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 – CORRIGENDUM. Data Policy 5:e23. https://doi.org/10.1017/dap.2023.21
Lnenicka M, Nikiforova A (2021) Transparency-by-design: What is the role of open data portals? Telematics Inform. https://doi.org/10.1016/j.tele.2021.101605
Lnenicka M, Luterek M, Nikiforova A (2022a) Benchmarking open data efforts through indices and rankings: Assessing development and contexts of use. Telematics Inform 66:101745. https://doi.org/10.1016/j.tele.2021.101745

Lnenicka M, Nikiforova A, Luterek M, Azeroual O, Ukpabi D, Valtenbergs V, Machova R (2022b) Transparency of open data ecosystems in smart cities: Definition and assessment of the maturity of transparency in 22 smart cities. Sustainable Cities Soc 82:103906. https://doi.org/10.1016/j.scs.2022.103906

Lnenicka M, Nikiforova A, Luterek M, Milic P, Rudmark D, Neumaier S, Santoro C, Flores C, Janssen C, M., Rodríguez Bolívar MP (2024) Identifying patterns and recommendations of and for sustainable open data initiatives: A benchmarking-driven analysis of open government data initiatives among European countries. Government Inform Q 41(1):101898. https://doi.org/10.1016/j.giq.2023.101898

Loenen BV, Zuiderwijk A, Vancauwenberghe G, Lopez-Pellicer FJ, Mulder I, Alexopoulos C, Magnussen R, Saddiqa M, Rosnay Dd, Crompvoets M, Polini J, Re A, B., Flores CC (2021) Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - eJournal eDemocracy Open Government. https://doi.org/10.29379/jedem.v13i2.644

Lu T, Zhang B, Ren K (2024) Privdata network: A privacy-preserving on-chain data factory and trading market. IEEE Trans Dependable Secur Comput. https://doi.org/10.1109/tdsc.2023.3284565

Lupi L, Antonini A (2020) Readapting Propp's character archetypes to explore the relational dimension of city data: A design-oriented approach [Readaptando los arquetipos de personajes propuestos por Propp para explorar la dimensión relacional de los datos de la ciudad: un enfoque de diseño]. Diseña 16:200–229. https://doi.org/10.7764/disena.16.200-229

Ma Y, Huang R (2024) Can the opening of public data improve the quality of urban employment? Evidence from China. https://doi.org/10.2139/ssrn.4857873

Mahmood S, Noorali AA, Manji A, Afzal N, Abbas S, Qamar JB, Siddiqi S, Hoodbhoy Z, Virani SS, Bhutta ZA, Samad Z (2023) Health data ecosystem in Pakistan: A multisectoral qualitative assessment of needs and opportunities. BMJ Open 13(9):e071616. https://doi.org/10.1136/bmjopen-2023-071616

Malthouse E, Buoye A, Line ND, El-Manstrly D, Dogru T, Kandampully J (2019) Beyond reciprocal: The role of platforms in diffusing data value across multiple stakeholders. J Service Manage. https://doi.org/10.1108/josm-12-2018-0381

McCord SE, Welty JL, Courtwright JL, Dillon C, Traynor A, Burnett SH, Courtright E, Fults G, Karl J, Zee VV, Webb J, N., Tweedie C (2021) Ten practical questions to improve data quality. Rangelands. https://doi.org/10.1016/j.rala.2021.07.006

McLeod M (2023) Managing Caribbean tourism data ecosystems. Curr Issues Tourism 27(2):1–17

McMahon S, Cusano J, Buttner C, Snyder S, Ast RS, Camerer K (2022) Evaluating efforts to address campus sexual violence: Developing a data ecosystem. J interpers Violence 37(23–24):NP23563–NP23586. https://doi.org/10.1177/08862605221078817

Molder EB, Schenkein SF, McConnell AE, Benedict KK, Straub CL (2022) Landsat data ecosystem case study: Actor perceptions of the use and value of Landsat. Front Environ Sci 9. https://doi.org/10.3389/fenvs.2021.805174

Möller J, Jankowski D, Lamm A, Hahn A (2022) Data management architecture for service-oriented maritime testbeds. IEEE Open J Intell Transp Syst. https://doi.org/10.1109/ojits.2022.3207235

Möller F, Jussen I, Springer V, Gieß A, Schweihoff J, Gelhaar J, Guggenberger T, Otto B (2024) Industrial data ecosystems and data spaces. Electron Markets. https://doi.org/10.1007/s12525-024-00724-0

Nadal S, Romero O, Abelló A, Vassiliadis P, Vansummeren S (2017) An integration-oriented ontology to govern evolution in big data ecosystems. EDBT/ICDT Workshops. https://doi.org/10.1016/j.is.2018.01.006

Näher A-F, Vorisek CN, Klopfenstein SAI, Lehne M, Thun S, Alsalamah S, Pujari S, Heider D, Ahrens W, Pigeot I, Marckmann G, Jenny MA, Renard BY, von Kleist M, Wieler LH, Balzer F, Grabenhenrich L (2023) Secondary data for global health digitalisation. Lancet Digit Health. https://doi.org/10.1016/s2589-7500(22)00195-9

Nam J, Jung Y, Kim J (2024) Understandings of the AI business ecosystem in South Korea: AI startups' perspective. Telecomm Policy 48(6):102763. https://doi.org/10.1016/j.telpol.2024.102763

Narayanan U, Veettil NP, Krishnankutty RT, Sunny LE, Paul V (2024) Mitigating privacy and security risks in the era of big data: A comprehensive framework of best practices and protocols. J Comput Sci. https://doi.org/10.3844/jcssp.2024.1121.1145

Nehra SS, Bansode SY (2023) India's open data initiatives: A way to knowledge discovery, transformation and to achieve sustainable development goals (SDGs) 2030. DESIDOC J Libr Inform Technol 43(1). https://doi.org/10.14429/djlit.43.01.18622

Nelson P, Low R, Kohl HR, Overoye D, Yang D, Huang X, Chellappan S, Azam FB, Carney RM, Falk M, Garriga J, Schelkin L, Boger R, Schwerin TG (2024) Globe observer: A case study in advancing earth system knowledge with AI-powered citizen science. Citiz Sci Theory Pract. https://doi.org/10.5334/cstp.747

Noardo F, Atkinson R, Bastin L, Masó J, Simonis I, Villar A, Voidrot-Martinez M-F, Zaborowski P (2024) Standards for data space building blocks. Remote Sens. https://doi.org/10.3390/rs16203824
Oldenburg J-R, Pussinen P (2024) Exploring data ecosystems in smart city service creation. Int J Innov Manag. https://doi.org/10.1142/s136391962440005x
Oliveira MIS, Lima GdFB, Lóscio BF (2019) Investigations into data ecosystems: A systematic mapping study. Knowl Inf Syst 61(2):589–630. https://doi.org/10.1007/s10115-018-1323-6
Olson G, Liu ZK (2023) Genomic materials design: Calculation of phase dynamics. Calphad. https://doi.org/10.1016/j.calphad.2023.102590
Ottaviano M, Beltrán-Jaunsarás ME, Terius-Padron JG, García-Betances RI, Gonzalez-Martinez S, Cea G, Vera C, Cabrera-Umpíerrez MF, Waldmeyer MTA (2019) Empowering citizens through perceptual sensing of urban environmental and health data following a participative citizen science approach. Sensors. https://doi.org/10.3390/s19132940
Otto B, Jarke M (2019) Designing a multi-sided data platform: Findings from the International Data Spaces case. Electron Markets 29(4):561–580. https://doi.org/10.1007/s12525-019-00362-x
Otto B, Hompel T, M., Wrobel S (eds) (2022) Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5
Palojoki S, Lehtonen L, Vuokko R (2024) Semantic interoperability of electronic health records: Systematic review of alternative approaches for enhancing patient information availability. JMIR Med Inf. https://doi.org/10.2196/53535
Palviainen M, Soininen J, Arnold M (2024) Requirements for water data ecosystems: Results from a business ecosystem case study. Environ Syst Decisions. https://doi.org/10.1007/s10669-024-09988-5
Parker G, Van Alstyne M, Jiang X (2016) Platform ecosystems: How developers invert the firm. SSRN Electron J 41. https://doi.org/10.2139/ssrn.2861574
Parliament and Council of the European Union (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). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868
Parliament and Council of the European Union (2023) Regulation (EU) 2023/2854 of the European parliament and of the council of 18 October 2023 on European data governance and amending regulation (EU) 2018/1724 (data act). https://eur-lex.europa.eu/eli/reg/2023/2854
Pascalis D (2022) The journey to open finance: Learning from the open banking movement. Eur Bus Law Rev. https://doi.org/10.54648/eulr2022018
Pellizzoni L, Falavigna A (2024) Connecting verified databases with clinical practice and the patient's experience through omnichannel communication. Int J Med Informatics 192:105639. https://doi.org/10.1016/j.ijmedinf.2024.105639
Pohlmeyer F, Möbitz C, Gries T (2024) A sovereign and interoperable data ecosystem for an eco-efficient nonwovens industry. Sustainability. https://doi.org/10.3390/su162310735
Pollock R (2011) Building the open data ecosystem. Open Knowl Foundation Blog. https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/
Pratt N, Camacho X, Vajdic C, Degenhardt L, Laba T, Hillen J, Etherton-Beer C, Preen D, Jorm LR, Donnolley N, Havard A, Pearson S-A (2022) The medicines intelligence centre of research excellence: Co-creating real-world evidence to support the evidentiary needs of Australian medicines regulators and payers. Int J Popul Data Sci. https://doi.org/10.23889/ijpds.v6i1.1726
Prinsloo P, Khalil M, Slade S (2023) Learning analytics as data ecology: A tentative proposal. J Comput High Educ. https://doi.org/10.1007/s12528-023-09355-4
Rajagopal K (2023) Personal learning environments as socio-technical systems: Does decentralised data finally give us the right balance? Revista de Educacíon Distancia (RED) 23(71). https://doi.org/10.6018/red.526851
Ramalli E, Pernici B (2023) Challenges of a data ecosystem for scientific data. Data Knowl Eng 148:102236. https://doi.org/10.1016/j.datak.2023.102236
Reggi L, Dawes SS (2022) Creating open government data ecosystems: Network relations among governments, user communities, NGOs and the media. Government Inform Q 39(2):101675. https://doi.org/10.1016/j.giq.2022.101675
Parliament and Council of the European Union (2022) Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance). http://data.europa.eu/eli/reg/2022/1925/oj/eng

Reiberg A, Niebel C, Kraemer P (2022) *What Is a Data Space?* Gaia-X Hub Germany (White Paper 1/2022). Retrieved from https://gaia-x-hub.de/wp-content/uploads/2023/11/GX-White-Paper-Data-Space.pdf

Rietveld J, Schilling MA (2021) Platform competition: A systematic and interdisciplinary review of the literature. J Manag 47(6):1528–1563. https://doi.org/10.1177/0149206320969791

Rix C, Stein H, Chen Q, Frank J, Maass W (2021) Conceptualizing data ecosystems for industrial food production. In: 2021 IEEE 23rd Conference on Business Informatics (CBI) (Vol. 01, pp. 201–210). https://doi.org/10.1109/CBI52690.2021.00031

Rojas-García J, Elias-Giordano C, Nallusamy S, Quiroz-Flores J (2024) Enhancement of the distribution process on light logistics SMEs in times post-pandemic COVID-19 with Ukraine-Russia conflict by lean logistics and big data. Social Sci Humanit Open. https://doi.org/10.1016/j.ssaho.2024.100945

Rong K, Luo Y (2023) Toward born sharing: The sharing economy evolution enabled by the digital ecosystems. Technological Forecast Social Change. https://doi.org/10.1016/j.techfore.2023.122776

Ruijer E, Dingelstad J, Meijer A (2023) Studying complex systems through design interventions probing open government data ecosystems in the Netherlands. Public Manage Rev 25(1):129–149. https://doi.org/10.1080/14719037.2021.1942533

Runeson P, Olsson T, Liňaker J (2021) Open data ecosystems — An empirical investigation into an emerging industry collaboration concept. J Syst Softw 182:111088. https://doi.org/10.1016/j.jss.2021.111088

Sakor A, Jozashoori S, Niazmand E, Rivas A, Bougiatiotis K, Aisopos F, Iglesias E, Rohde PD, Padiya T, Krithara A, Paliouras G, Vidal M-E (2023) Knowledge4covid-19: A semantic-based approach for constructing a COVID-19 related knowledge graph from various sources and analyzing treatments' toxicities. J Web Semant 75:100760. https://doi.org/10.1016/j.websem.2022.100760

Scassa T (2019) Ownership and control over publicly accessible platform data. Online Inf Rev. https://doi.org/10.1108/oir-02-2018-0053

Scheider S, Lauf F, Geller S, Möller F, Otto B (2023a) Exploring design elements of personal data markets. Electron Markets 33(1):28. https://doi.org/10.1007/s12525-023-00646-3

Scheider S, Lauf F, Möller F, Otto B (2023b) A reference system architecture with data sovereignty for human-centric data ecosystems. Bus Inform Syst Eng 65(5):577–595. https://doi.org/10.1007/s12599-023-00816-9

Schreieck M, Wiesche M, Krcmar H (2022) From product platform ecosystem to innovation platform ecosystem: An institutional perspective on the governance of ecosystem transformations. J Association Inform Syst 23(6):1354–1385

Schulze E-D, Beck E, Müller-Hohenstein K (2005) Plant ecology. Springer Science & Business Media

Seneviratne O, Adams K, McGuinness D (2023) Accountable bench-to-bedside data-sharing mechanism for researchers. ACM Trans Social Comput. https://doi.org/10.1145/3609486

Sestak M, Copot D (2023) Towards trusted data sharing and exchange in agro-food supply chains: Design principles for agricultural data spaces. Sustainability 15(18):13746. https://doi.org/10.3390/su151813746

Shadbolt N, Brett A, Chen M, Marion G, McKendrick IJ, Panovska-Griffiths J, Pellis L, Reeve R, Swallow B (2022) The challenges of data in future pandemics. Epidemics 40:100612. https://doi.org/10.1016/j.epidem.2022.100612

Shah SIH, Peristeras V, Magnisalis I (2021) Government big data ecosystem: Definitions, types of data, actors, and roles and the impact in public administrations. J Data Inform Qual 13(2):8:1–825. https://doi.org/10.1145/3425709

Shah SAH, Peristeras V, Magnisalis I (2024) A conceptual framework for the government big data ecosystem ('datagov.eco'). Data Knowl Eng. https://doi.org/10.1016/j.datak.2024.102348

Sharma V, Tripathi A, Mittal H (2022) Technological revolutions in smart farming: Current trends, challenges & future directions. Comput Electron Agric. https://doi.org/10.1016/j.compag.2022.107217

Skelly L, Chiware ERT (2022) African researchers do not think differently about open data. Front Res Metrics Analytics. https://doi.org/10.3389/frma.2022.950212

Sotirova-Valkova K (2024) A pilot conceptualisation of the data space ecosystems for cultural heritage. Math Educ Math. https://doi.org/10.55630/mem.2024.53.092-098

Styrin E, Luna-Reyes LF, Harrison TM (2017) Open data ecosystems: An international comparison. Transforming Government: People Process Policy. https://doi.org/10.1108/tg-01-2017-0006

Supinajaroen W, Van Loenen B, Korthals Altes W (2022) Open national CORS data ecosystems: A cross jurisdictional comparison*. Interdisciplinary Description Complex Syst 20(2):78–95. https://doi.org/10.7906/indecs.20.2.2

Susha I, van den Broek T, van Veenstra A-F, Lin˚aker J (2022) An ecosystem perspective on developing data collaboratives for addressing societal issues: The role of conveners. Government Inform Q 101763. https://doi.org/10.1016/j.giq.2022.101763
Taherdoost H (2024) A systematic review of big data innovations in smart grids. Results Eng. https://doi.org/10.1016/j.rineng.2024.102132
Tang J, Zhu H, Guo J (2022) Information diffusion between users in open data ecosystem: Modelling and simulation analysis. Math Probl Eng 2892274. https://doi.org/10.1155/2022/2892274
Thinyane M (2018) Towards informing human-centric ICT standardization for data-driven societies. J ICT Stand. https://doi.org/10.13052/jicts2245-800x.631
Toorajipour R, Oghazi P, Palmié M (2024) Data ecosystem business models: Value propositions and value capture with artificial intelligence of things. Int J Inf Manag. https://doi.org/10.1016/j.ijinfomgt.2024.102804
Touré V, Krauss P, Gnodtke K, Buchhorn J, Unni D, Horki P, Raisaro JL, Kalt K, Teixeira D, Crameri K, Osterle S (2023) FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network. Sci Data 10(1):127. https://doi.org/10.1038/s41597-023-02028-y
van Donge W, Bharosa N, Janssen MFWHA (2022) Data-driven government: Cross-case comparison of data stewardship in data ecosystems. Government Inform Q 39(2):101642. https://doi.org/10.1016/j.giq.2021.101642
van Waerdt D, P. J. V (2023) 'Everything the data touches is our kingdom': Market power of 'data ecosystems'. World Competition, 46(1):65-98
Vandercruysse L, D'Hauwers R, Bourgeus A (2023) Citizen-centric personal data storage in the public sector: An exploration of opportunities, challenges, and preconditions. J Innov Manage. https://doi.org/10.24840/2183-0606_011.003_0002
Vasto-Terrientes LD (2023) Experience: Data management for delivering covid-19 relief in Panama. J Data Inform Qual. https://doi.org/10.1145/3623511
Velmovitsky P, Kirolos M, Alencar P, Leatherdale ST, Cowan D, Morita P (2023) Leveraging mHealth technologies for public health. JMIR Public Health Surveillance. https://doi.org/10.2196/49719
Verstraete M, Fierens M, Verbrugge S, Colle D (2023) Assessing the economic viability of data intermediation services as defined in the DGA
Villarreal V, Munõz L, González J, Fontecha J, Dobrescu C, Nielsen M, Concepcion D, Rodríguez M (2024) A methodological approach for data standardization and management of open data portals for scientific research groups: A case study on mobile and ubiquitous ecosystems. Front Comput Sci. https://doi.org/10.3389/fcomp.2024.1420709
Wainwright T, Huber F, Stöckmann C, Kraus S (2023) Open data platforms for transformational entrepreneurship: Inclusion and exclusion mechanisms. Int J Inf Manag 72:102664. https://doi.org/10.1016/j.ijinfomgt.2023.102664
Wang R, Xu CK, Wu X (2024) Open government data (OGD) as a catalyst for smart city development: Empirical evidence from Chinese cities. Government Inform Q. https://doi.org/10.1016/j.giq.2024.101983
Ward T, Aykol M, Blaiszik B, Foster IT, Meredig B, Saal J, Suram S (2018) Strategies for accelerating the adoption of materials informatics. MRS Bull. https://doi.org/10.1557/mrs.2018.204
Weber R (2012) Evaluating and developing theories in the information systems discipline. J AIS 13. https://doi.org/10.17705/1jais.00284
Weber L, Brinkhaus J, Dumss S, Henrich V, Hoffmann F, Ristow G, Schickling C, Trautner T, Grafinger M, Weigold M, Bleicher F (2022) Euprogigant resilience approach: A concept for strengthening resilience in the manufacturing industry on the shop floor. Procedia CIRP. https://doi.org/10.1016/j.procir.2022.05.022
Wendelborn C, Anger M, Schickhardt C (2023) What is data stewardship? Towards a comprehensive understanding. J Biomed Inform 140:104337. https://doi.org/10.1016/j.jbi.2023.104337
Wilkinson R, Mleczko M, Brewin RJW, Gaston K, Mueller M, Shutler J, Yan X, Anderson K (2023) Environmental impacts of earth observation data in the constellation and cloud computing era. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2023.168584
Williams M, Moser T (2019) The art of coding and thematic exploration in qualitative research. Int Manage Rev 15(1):45–55
Williams SR, Agrahari Baniya A, Islam MS, Murphy K (2023) A data ecosystem for orchard research and early fruit traceability. Horticulturae 9(9):1013. https://doi.org/10.3390/horticulturae9091013
Winters P, Hogue E, Steiner M, Keramati C, Rokhideh M, Eissler S (2022) Facilitating data use for decision-making: 50x2030's approach. Stat J IAOS. https://doi.org/10.3233/sji-210908

Woods O, Bunnell T, Kong L (2023) The state-led platformisation of financial services: Frictionless ecosystems and an expansive logic of smartness in Singapore. Geoforum 146:103849. https://doi.org/10.1016/j.geoforum.2023.103849

Xu J, Hong N, Xu Z, Zhao Z, Wu C, Kuang K, Wang J, Zhu M, Zhou J, Ren K, Yang X, Lu C, Pei J, Shum H (2023) Data-driven learning for data rights, data pricing, and privacy computing. Engineering 25:66–76. https://doi.org/10.1016/j.eng.2022.12.008

Ye Y, Zhou A, Shi X, Huang C (2023) A seed model for constructing the data factor market: Evidence from Guiyang global big data exchange (GBDEX) in China. J Digit Econ. https://doi.org/10.1016/j.jdec.2023.03.002

Zhang M, Gable G, Rai A (2016) Toward principles of construct clarity: Exploring the usefulness of facet theory in guiding conceptualization. Australasian J Inform Syst, 20 https://doi.org/10.3127/ajis.v20i0.1123

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Maarten de Mildt[1,2] · Sander Van Damme[1,2] · Sofie Verbrugge[1,2] · Didier Colle[1,2]

✉ Maarten de Mildt
maarten.demildt@ugent.be

✉ Sander Van Damme
sander.vandamme@ugent.be

Sofie Verbrugge
sofie.verbrugge@ugent.be

Didier Colle
didier.colle@ugent.be

[1] Ghent University, Ghent, Belgium

[2] Imec, Leuven, Belgium