Telecommunications Policy 49 (2025) 102910

Contents lists available at ScienceDirect

Telecommunications Policy

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

A market analysis on data ecosystem initiators and their value propositions in different ecosystems

Maarten de Mildt *, Sofie Verbrugge, Didier Colle

IDLab, Department of Information Technology at Ghent University - imec, Technologiepark 126, Ghent, Belgium

ARTICLE INFO

Dataset link: https://github.com/TechnoEconomics/DataEcosystemMarketAnalysis

Keywords:
Data-driven business model
Data ecosystem
Value proposition
Market analysis

ABSTRACT

The concept of data ecosystems is proposed as a solution to organise and manage increasingly complex data processes. A common narrative found in literature is that, using an ecosystem approach, an increased amount of data will be available, of higher quality, and this will lead to an increase in innovation. Even though these benefits seem promising at first, data ecosystems are not standard practice yet. A critical look at the way the market is implementing this concept is needed. A market analysis was performed on the different services that initiate and support data ecosystems. Three main ecosystem types, based on their scope, were identified: (1) single-organisation ecosystems, (2) application domain specific ecosystems, and (3) cross-domain ecosystems. Challenging these benefits of data ecosystems, this market analysis revealed that achieving the innovation potential and actual data reuse remains a challenge, especially for more broadly scoped ecosystems. Consequently, a successful data ecosystem implementation will require a critical assessment of its data reuse and innovation potential.

1. Introduction

The demand for data and accompanying services is ever-increasing, which poses several challenges in managing data driven systems. This demand also leads to an assumed large value of data (Koutroumpis, Leiponen, & Thomas, 2020). However, Koutroumpis et al. (2020) state that data is not shared or traded openly, nor transparently, on a large scale. In response, an evolution towards a more open or distributed approaches to data management has been proposed (European Commission, 2020), often referred to as Data Ecosystems (S. Oliveira, Barros Lima, & Farias Lóscio, 2019). This evolution is one of the responses to a larger (r)evolution on the Web. More data is being generated, more data is being used, processes increase in complexity. Accordingly, data has become an important asset in many business models (Bulger, Taylor, & Schroeder, 2014); (Hartmann, Zaki, Feldmann, & Neely, 2016), which also increased general awareness on data usage. However, this has made data a necessary evil for many organisations for which data in itself is not the service they offer (e.g. storing contact information of clients). At the same time, organisations who control large amounts of data become increasingly powerful. Both these data management difficulties and unbalanced power distributions are reasons to believe that the way data is handled, must change.

As part of the European Commission's European strategy for data (European Commission, 2024b), an abundance of recent policy regulations focus on this general awareness on how data is used and its importance as a strategic asset. These include the Data Governance Act (DGA) (European Parliament and the Council, 2022b) and the Data Act (DA) (European Parliament and the Council, 2023). Their main goal of this strategy is to set up a single data market, where data can freely flow for innovation. Additionally, in this strategy they are eager to name several economic and societal benefits like improved healthcare, generating new services,

* Correspondence to: Techonologiepark Zwijnaarde 126, Ghent, Belgium.
E-mail address: maarten.demildt@ugent.be (M. de Mildt).

https://doi.org/10.1016/j.telpol.2025.102910
Received 29 September 2023; Received in revised form 3 January 2025; Accepted 21 January 2025
Available online 6 February 2025

reducing costs of public services. Specifically, the Commission positions these regulations as to lead to increase (trusted) data sharing, and overcome obstacles to the better use and especially reuse of data. However, practical problems with these regulations have been the subject of debate. von Ditfurth and Lienemann (2022) argue that the DGA aggravates the obstacles for services that act as intermediaries in data sharing and data reuse, rather than alleviate them. Additionally, the business viability of being an intermediation service under DGA is questioned (Verstraete, Fierens, Verbrugge, & Colle, 2023), stirring the debate that not complying with regulations actually provides a market advantage. Both this regulatory uncertainty and its rapid evolution, in combination with data ownership debates, lead to flaws in current data sharing systems (Koutroumpis et al., 2020). Adding up these practical issues, gives reason to believe the anticipated innovation is not being achieved.

At the same time, such regulations further promote the push towards increasingly open and distributed approached to better handle data. Specifically, the notion of sharing data (in data ecosystems) to make siloed data (more) usable, is put forward as a possible solution to data (re)use inefficiencies (Susha, van den Broek, van Veenstra, & Linåker, 2023). Sharing data between organisations is believed to foster innovation and help tackle societal issues. Open (government) data initiatives being a common example of data ecosystems being implemented as a way to share data outside organisational boundaries (Runeson, Olsson, & Linåker, 2021). Against a similar data sharing backdrop, initiatives that try to share data more efficiently pop up in Europe, of which some are even expanding outside of Europe (*Common European Data Spaces*, 2024); (*International Data Spaces Association*, 2024b); Farrell et al. (2023). By aiming to create this single data market, based on common data infrastructures and governance frameworks, these legislation-backed initiatives typically intend to break up data silos and facilitate multilateral data sharing.

Even though these goals to better handle data seem promising, for many organisations it seems to remain a challenge to unlock the potential of (their) data. This paper aims to look at how the market of data sharing is currently adapting to them. Therefore, it is not an in-depth critique on the content of the regulations itself, but rather a study into how these translate into practice. The organisations and companies that implement, in varying ways, data ecosystem services will be considered as the "market of data ecosystems initiators" in this paper. More information on this market analysis set-up is given in Section 3. The hypothesis tackled in this paper, is that there are different ways to set up data sharing, and this influences the innovation potential. It is believed that even though many companies and organisations position themselves as data ecosystems, in which data is sharable with the goal of reusing it, they do not actually reach this goal. Therefore, tackling the following research question:

RQ: How are data ecosystems implemented in reality by the market of data ecosystem services, and do they achieve their envisioned benefits?

The remainder of the paper is structured as follows. Section 2 addresses data ecosystem literature in order to establish a definition and relevant input used in later analysis, like characteristics and benefits of data ecosystems. Next, the data gathered and method used to perform the analysis is discussed in Section 3. The market analysis is done in Section 4, which offers insights on what the dataset reveals about data ecosystem types and benefits, and several detailed examples from the dataset are discussed. What follows is a discussion in Section 5, in which theoretical, practical and policy implications of the findings are provided. Section 6 concludes the paper by emphasising the implications these insights have on evaluating data ecosystems as a successful data management approach.

2. Background on data ecosystems

To study the market of data ecosystem initiators and their services, a range of background info on data ecosystems is needed. First, how the concept of data ecosystems is defined, is outlined, to know what organisations and solutions are in scope. Secondly, some more background on the relevant policies in light of the European strategy for data allows for a critical analysis of their envisioned benefits. Consequently, the theoretical benefits of data ecosystems are appropriate to align with literature. In order to classify the market of data ecosystem initiators and the services they offer, typical characteristics and services of data ecosystems are needed. A reflection on what literature commonly states on the link between these characteristics and perceived benefits finally reveals expected benefits of each different ecosystem.

2.1. Data ecosystem definition

Data ecosystems as a facilitator of managing data-driven processes propose a solution to the problems mentioned in Section 1, like getting data out of silos. In data ecosystems, typically data, contributions to it, and management of it, is distributed over all involved parties. Given the concept is not a novel one, different definitions and perspectives are found in literature. An overview of the most common ones is listed in Table 1. These were gathered from technical, socio-technic and economic literature. Although reflecting differing views based on different perspectives, this table shows similarities are to be found. As these definitions show, network effects are essential to ecosystems. Accordingly, data ecosystems are considered a type of digital ecosystems (Lis & Otto, 2021). They defined these digital ecosystems as a collaborative network of multiple actors using the correlating technical infrastructure. Such ecosystems in its turn are considered part of digital platforms in de Reuver, Sørensen, and Basole (2018), defining digital platforms as both "An extensible codebase to which complementary third-party modules can be added" (technical view) and "Technical elements and associated organisational processes and standards" (sociotechnical view). Again, emphasising that multiple perspectives must be taken into account when talking about data ecosystems.

Table 1
Several definitions of data ecosystems in related work, including the working definition considered in this paper.

Definition	Source
"Data ecosystems are socio-technical complex **networks in which actors interact and collaborate** with each other to find, archive, publish, consume, or reuse data as well as to **foster innovation, create value**, and support new businesses."	S. Oliveira et al. (2019)
"An ecosystem has data cycles, in which **intermediate consumers of data** such as builders of apps and data wranglers may **share** back their cleaned, integrated, and packaged **data into the ecosystem in a reusable way**. This cleaned and integrated data is often **more valuable** than the original source."	S. Oliveira et al. (2019); Rufus Pollock (2011)
"It is made up of many **actors and small organisational structures** that should recognise **data** like the **raw material** that is in a cycle and is capable of feeding the ecosystem, providing **benefits to all parties**."	S. Oliveira et al. (2019); Ubaldi (2013)
"A set of **interacting actors** that directly or indirectly **consume, produce, provide data** and other related resources"	Lis and Otto (2021)
A set of actors that facilitate the production, consumption, management and exchange of data, which creates value for all actors involved	Working definition

Working definition. Using the common elements in several of these definitions, data ecosystems are considered here as "A set of actors that facilitate the production, consumption, management and exchange of data, which creates value for all actors involved". While this definition contains useful elements, it is still a broad one. As mentioned in the introduction, this paper will delve deeper into specifics of data ecosystems, and show that this broadness results in varying implementations and ecosystem types. When referring to the concept of data ecosystems for the remainder of the paper, this definition can be assumed.

2.2. European strategy for data

As mentioned in the introduction, the European strategy for data is a major driver of the development of data ecosystems. Specifically, they drive the creation of what is called Common European Data Spaces (*Common European Data Spaces*, 2024), which must make more data available, and increase the reuse of data. Given this goal of data reuse is essential to the success of data ecosystems, a deeper look into this strategy, and its relevant policies is appropriate.

These Common European Data Spaces have a few key characteristics. They are considered open for participation to all, this includes both organisations and individuals. At the same time, these data spaces should guarantee security, and be privacy-preserving, through its infrastructure that must allow actors to pool, access, share, process and use data. Making data accessible and useable for all actors is another requirement. This includes a fair, transparent and non-discriminatory governance framework for data access. EU law must naturally be respected, among which personal data and consumer protection, as well as competition law. Additionally, data holders should be able to grant access to or to share data through such data spaces. Finally, data holders must not only be allowed to, but even given the means to be empowered to make their data available for reuse (potentially in return for compensation). All of this must lead to the creation of a single data market, in which data is reused for innovative purposes. Given these requirements, especially regarding personal data, consumer protection and competition law, several relevant policies come into play that will be discussed here shortly in relation to the goal of more available data for (re)use: General Data Protection Regulation (GDPR) (European Parliament and the Council, 2016), the Data Act (DA) (European Parliament and the Council, 2023), the Data Governance Act (DGA) (European Parliament and the Council, 2022b), and the Digital Markets Act (DMA) (European Parliament and the Council, 2022a).

Perhaps the longest standing policy of the General Data Protection Regulation, is relevant whenever personal data is involved. Data ecosystems are therefore no exception. However, in the context of having data available and reusable, it has been critiqued before. Custers and Uršič (2016) argue that, in the field of Big Data, this may limit the social and economic benefits that could result from insights into this data. They provide a nuanced picture, pleading both less restrictive, and more restrictive handling of data on data subjects, depending on the awareness and intentions of the data reuse. According to them, data reuse in line with the awareness and intentions of data subjects should be approached less restrictively (they give the example of assuming informed consent). Data reuse where data subject's rights are more difficult to exercise, additional protection is to be considered (they provide the example of requiring explicit consent). In light of data ecosystems, where actors are typically actively participating in data sharing, such consents are highly relevant, and possibly needed for data sharing and (re)use to reach its potential (Van Ooijen & Vrabec, 2019). Custers and Bachlechner (2017) also note that data reuse of third-party data is rarely happening. Accordingly, this critique on the GDPR applies here as well, supporting the question if data ecosystems achieve their envisioned benefits, and can overcome these data reuse barriers, or are limited by current regulation. Crepax, Gaur, and da Rosa Lazarotto (2023) acknowledge a similar contradiction between the protection of personal rights on the one hand, and increasing data availability and reusability. They argue, however, that the Data Act's focus on technical interoperability should lead to improvement of accessibility. Where accessibility might increase, the DA seems to fail from achieving an increase in actual reuse of data due to its lack of principles to make data sharing safer and more fair (da Rosa Lazarotto, 2022). In their study on the DMA, and its relation to the other regulations mentioned, Schweitzer and Metzger (2023) observe similar issues for data access and its use in digital ecosystems (of which data ecosystems can be considered a subtype). The DMA might not limit ecosystem orchestrators (which is also a type of data ecosystem actor among the organisations this study covers Susha et al. (2023)) to exploit their position in the ecosystem. This given the

Table 2
Summary of the characteristics used in the analysis to determine how an ecosystem is set up.

Characteristic	Values	Criteria
Openness	Open	New and existing actors can participate throughout the ecosystem's operation
	Closed	Only actors present at the establishment of the ecosystem can participate
Data access	Siloed	Data usage is locked-in to the functionalities of specific applications
	Application independence	Data usage is based on (open) standards and protocols that allow actors to share data outside of specific applications
Initiator goal	Data sharing	The goal is to maximise data reuse and accompanying value created with said data
	Privacy protection	Data is shared under strict policies, with public values in mind

unclarity whether they are to be considered a 'gatekeeper', as the DMA prescribes. Therefore, orchestrators could limit data (re)use by refusing to share data with certain actors, for example.

What is apparent in all these studies is the uncertainty and critique on whether data reuse can be increased, and data made more (readily) available. Therefore, if data ecosystems are to solve data silo problems, and be a means for data sharing, to increase data reuse and availability, there is reason to investigate if they face similar problems.

2.3. Benefits and innovation in data ecosystems

Having established the concept, a look at the perceived benefits of data ecosystems in literature is needed to get an overview of what is known on if they achieve the envisioned goals of the European Commission. These benefits of data ecosystems have been widely discussed from different perspectives (S. Oliveira et al., 2019); (Runeson et al., 2021); (Susha et al., 2023). In these sources, among others, an ecosystem approach is expected to lead to more control and transparency for individual actors, the elimination of duplicate data processes, and generally more data sharing, carrying innovation potential. This leads to three general expected benefits, found in the mentioned literature, considered in this paper:

- **More available data**
- **Improved data quality**
- **Increased innovation**

Regarding the first expected benefit listed, more available data, (S. Oliveira et al., 2019) state that Data Ecosystems should ease the consumption of data. Actors who are not able to consume or produce data themselves, can make use of intermediation services in Data Ecosystems to do so. Smaller actors who currently miss out due to siloed data, will be able to partake because of expected lower data-related costs of knowledge and resources (Runeson et al., 2021). Susha et al. (2023) Focus on the collaborative nature of sharing across different sectors, making untapped sources of data available. All of these arguments essentially increase the amount of data sources, and generally data available for use. On improved data quality, feedback mechanisms in the ecosystem must improve the correctness and quality of data (S. Oliveira et al., 2019). Runeson et al. (2021) add to this view that quality of data also increases since more data sources will be connected, from which comparisons of, and additions to, data can be done. Finally, data ecosystems should enable new business opportunities like cooperations with other ecosystem actors by using data and accompanying services present in the ecosystem (S. Oliveira et al., 2019). Eventually, such new business opportunities are even expected to create new markets for data sharing altogether (Runeson et al., 2021). The goals of these new business opportunities mentioned are to enable innovation and value creation, which we summarise here as increased innovation. Susha et al. (2023) mention the expectations of data collaboration to also contribute to social innovation. In summary, economic and societal growth are expected from data ecosystems. The ambitious nature of these expected benefits strengthen the need for a critical look at the market of data ecosystem initiators.

2.4. Ecosystem characteristics: Openness, data access & initiator goal

Next, how related work characterises different data ecosystem implementations is relevant to systematically categorise data ecosystem implementations. On this matter, several characteristics can be found in literature, of which three general ones are considered relevant here: Openness, Data access and Initiator Goal, defined in the following sections. These characteristics allow determining how ecosystems differ from one another in the market analysis phase of Section 4.1. A summary of the criteria used for this analysis is given in Table 2. For each characteristic the criteria is given to which it should adhere to have which of its values. Each criterion is then used in later Section 4.1 to characterise each data ecosystem in the dataset discussed in Section 3.

Openness of the ecosystem represents the possibility for new actors to participate in the ecosystem who are not part of predefined processes, or known when the ecosystem was set up. Organisations thus have to make the decision whether they partake in, or set up, an open ecosystem. Mukhopadhyay and Bouwman (2019) describe membership openness as both being open or closed to multiple (external) individuals and firms. Open ecosystems are also often represented by open data initiatives (Lis & Otto, 2021), which let any actor contribute and use data freely. Zuiderwijk, Janssen, and Davis (2014) describe the open data ecosystem as loosely interconnected participants in self-organising and autonomous ecosystems. An open ecosystem thus allows actors to enter (and leave) throughout the lifetime of the ecosystem, albeit by meeting certain objective requirements. S. Oliveira et al. (2019) describes

such requirements as the institutional context of the ecosystem in which the actors' behaviour is restrained, a view confirmed by Mukhopadhyay and Bouwman (2019). The alternative being a closed ecosystem. Only actors present at the time of establishing the ecosystem are then allowed to participate, and there being no open possibility to apply for participation. In this case, if certain actors leave the ecosystem, the ecosystem would typically cease to exist.

Data access defines dependency of actors on certain applications to access, share and reuse data. An organisation has to decide if they partake in, or set up an ecosystem in which all actors to access data and/or exercise certain rights on the data that flows in the ecosystem. Lis and Otto (2021) describe this as Data Ownership and Decisions Rights, which respectively represent who controls data and who can make decisions on data sources. The concept of decision rights is also described by Mukhopadhyay and Bouwman (2019) in regard to platform leaders and its complementers. The two types of data access distilled for the remainder of the paper are siloed and application independence access. If data is siloed (Krämer, Shekhar, & Hofmann, 2022), it is deliberately kept within an application's functioning, meaning there is no way to share data using other applications. On the other hand, if actors are able to act independent of applications, an application independent ecosystem policy is created. This fits the description of Personal Data Ecosystems in S. Oliveira et al. (2019), in which all actors control their own data. At this point, actors should be able to extract data, independent of the tools they used to generate the data, mitigating vendor lock-in.

Finally, the **initiator goal** shows the goal the initiators envision by setting up an ecosystem. Two movements can be distinguished here: creating additional value and privacy protection. Either the goal is to share data and collaborate in order to create economic value (Krämer et al., 2022); (Lis & Otto, 2021), or to focus on protecting the privacy of the actors. When value creation is envisioned, an ecosystem that facilitates business partnerships can be a typical example (Baars et al., 2021). Think of connecting manufacturers of products, and retailers who could pool their market insights to both get better deals, and target specific customer segments. On the other hand, privacy protection is typically centred around public values, rather than economic gains.

2.5. Link between benefits and characteristics

Next to benefits and characteristics, literature provides the input for deducting an expected link between data ecosystem characteristics (Section 2.4) and benefits (Section 2.3). Typically, use cases in data ecosystems with specific characteristics are analysed to deduct general benefits. For example, if a benefit found is that more data is contributed in ecosystems, and the case studies used were open ecosystems, then a benefit that can be linked to open ecosystems, is increased data availability. The dominant benefit-characteristic links found as such are discussed in the following paragraphs. This allows for examining if expected theoretical benefits are found in reality, after performing the market analysis, in Section 4.1.

Open ecosystems improve data quality. An open ecosystem leads to an improvement in data quality (Susha et al., 2023); (Runeson et al., 2021); (S. Oliveira et al., 2019); (Mn & Luna-Reyes, 2017); (Hardjono & Pentland, 2019). Three, intertwined, reasons are discussed here shortly: verifiability, accountability and up-to-date data. First, when actors actively contribute data about themselves, or relating to others, this introduces feedback loops. Actors have the possibility to verify data, question data and trace it back to its source. Runeson et al. (2021) found that the expectations in Open Data Ecosystems are indeed that annotation of data by multiple parties (thus not only by the data provider) must increase quality. Therefore, incorrect data can be called out, and the responsible parties are present in the ecosystem to act on it. In the case of AmDex, an example of an open data market, this accountability manifests itself in data quality assurance mechanisms (Susha et al., 2023). The same goes for outdated data. If a customer notices that a company still contacts them based on an old address, an open ecosystem allows them to spread the message of having changed addresses. One such example is found in the context of data cooperatives (Hardjono & Pentland, 2019), where each individual can maintain their own data, essentially leading to more verified and up-to-date data, which is considered higher quality data.

Application independence leads to more data. If data access is distributed over the ecosystem, it is believed to lead to more available data (Susha et al., 2023); (S. Oliveira et al., 2019); (Hardjono & Pentland, 2019); (Mukhopadhyay & Bouwman, 2019). Again, a few causes are given: increased transparency, standardisation and more democratic processes. When actors all contribute data about themselves, at free will, they can get insights into who is using which data for which purpose. This more transparent way of sharing data should lead to an increased willingness to share data. In the context of data cooperatives, for example, this active contribution of data by members is believed to lead to more data being contributed (Hardjono & Pentland, 2019). In return, for this specific case, the cooperative provides transparency in the form of a privacy preserving algorithm. Transparency is then in the fact that the algorithm can be verified to follow certain principles, like never copying data. In order to be able to share data, actors need to speak a common language in the form of standards. The lack of standards is named as a common reason why data is not shared (Runeson et al., 2021). While wanting to share data leads to more standardisation, more standardisation also leads to more data being shareable. Susha et al. (2023) showing that use cases like AmDex firmly believe in this vision, mentioning that standardisation could remove the need for bilateral agreements, thus increasing the amount of data that is shared. Finally, if applications do not limit their users, (more) people are empowered to contribute (more) data.

Data sharing brings forth innovation. Thirdly, data sharing must lead to innovation (S. Oliveira et al., 2019); (Susha et al., 2023); (Krämer et al., 2022); (Baars et al., 2021). A shift in competition, the possibility to co-produce value, and funds that can be re-allocated to different activities explain this benefit. When breaking down silo's, sharing data, there is less competitive advantage of having done extensive data mining and keeping your data hidden away from external parties. This might cause a shift towards the more innovative services to win over the market (Runeson et al., 2021). Even more so, companies who remain indifferent to partaking in sharing data could fall behind due to missing out on the possibility to co-produce value. Runeson et al. (2021) confirm this view in that they found that it is not the data itself that is believed to be valuable, but rather the collaboration resulting in

co-production of value. Finally, if, for example, data mining is not a key activity for organisations, they can re-allocate these funds. As a result of data transactions, the ecosystem's functioning as a whole could even be improved to tackle such inefficiencies by making it possible to abandon certain data services for alternatives (S. Oliveira et al., 2019).

As mentioned, most studies on data ecosystems that deduct these benefits of data ecosystems, study use cases. However, these cases are typically in pre-mature stages or success stories of data ecosystems. The perceived benefits are therefore often times more expected than actually perceived benefits. This raises the question if these statements can actually be generalised for data ecosystems, or are a result of, for example, a success story of an initial pilot project. A major concern is if the studied ecosystems actually scale up to more use cases where data is reused. Next to this, on the level of each positive benefit-characteristic link, a reflection must be made on potential risks. An open ecosystem that should improve data quality, also opens itself up to more potentially incorrect or even malicious data. While the mechanisms mentioned above, like having feedback loops, would solve this issue, these require additional governance mechanisms (e.g. consensus algorithms as used in blockchain to resolve data irregularities (Lashkari & Musilek, 2021)). This in turn possibly increases complexity and costs, which cannot be denied when denoting these benefits. The same goes for application independence and data sharing. Implementing standards, for example, and transacting with others to share data is again a possibly complex and costly operation. A critical look at these benefits is therefore needed.

2.6. Data ecosystem business models and services

The final input needed from literature to perform the market study is that of the different types of services and business models offered by ecosystem initiators. These are the actor(s) in the ecosystem that set-up and facilitate the development of the ecosystem. The, possibly indirect, beneficiaries should be all other ecosystem actors. These are needed to make a classification of each organisation in the dataset's own (marketing) description of their services. This then allows for a more consistent overview of the common services in the dataset. An overview of common services offered in data-driven and big data business model literature, is shown in Table 3. This table includes both the description of the service and the literature sources in which it was found, and these descriptions determine the services offered by each initiator in the dataset.

Short descriptions of each service can be found in the table, however some additional reflection on what these services mean in a data ecosystem context is appropriate. Some common IT services like consultancy, infrastructure, matchmaking and Software-as-a-Service are also common in data ecosystems. These services are given a data-related context, but do not defer much from their non-data-related counterpart. More interesting, however, are services more specific to data ecosystems. These include data standardisation, which is needed to share data with other actors in the ecosystem and have this data be usable for all actors. Such services transform data from unstructured or differently formatted data into specific data formats. When sharing this data, access granting mechanism services provide a way to make your data useable while remaining in control of whom uses it. Through an access granting service, an actor could, for example, be provided with an overview of all current actors in the ecosystem that use (some of) their data. This then allows them to grant specific actors in the ecosystem access to their data. Additionally, to provide legitimacy to such access grants, governance tools and clearinghouse services can instate trust in these data exchanges. Governance tools could allow actors to instate certain policies to their liking (e.g. only share medical data with licensed professionals) and a clearinghouse in its turn verifies if said policies are respected upon a data exchange request. While these services allow data to be exchangeable in trusted fashion, a Data-as-a-service provider adds additional value by making data findable. By aggregating data from different sources, and providing for example a meta-layer to make data query-able, they increase the overall discoverability of data. An important takeaway here is that these data ecosystem services complement one another, each adding additional value to data exchanges in ecosystems. At the end of such a chain of complementary services could then be a data analytics service, that in its turn makes the data insightful to serve a specific purpose (in a use case).

3. Data & methods

The literature review of Section 2 described several properties of data ecosystems, that are used as input in the further analyses. This knowledge now allows selecting which market players are in-scope, categorise these market players to draw relevant conclusions, and build an appropriate method for analysis. Having investigated the market of data ecosystem initiators, a dataset of all the solutions found was put together. In the following paragraphs, a description is given of this dataset, and how it was analysed methodologically.

3.1. Market data gathering

From exploring the market of data ecosystems, a dataset was constructed of a little over 100 organisations, companies, and initiatives that were considered initiators of ecosystems. Several criteria had to be met by each of them, and follow the working definition of data ecosystems in Section 2:

- Must be an organisation, company or other organised initiative, not an individual
- The provided service(s) must facilitate interaction between ecosystem actors
- Data must be a key resource of the ecosystem
- The selected organisation's service(s) must initiate the ecosystems, meaning they (help) set up the ecosystem
- Must have an intermediary role (additional roles allowed) in the value chains of their ecosystems

Table 3
Overview of common data services that apply to data ecosystems.

Service	Description	Sources
Consultancy	General advisory services, provisioning of expertise knowledge on data-driven processes.	Bulger et al. (2014); Susha et al. (2023)
Infrastructure	Provisioning of hardware components and their management, as well as dedicated (software) technology integrations onto such hardware components.	Bulger et al. (2014); Susha et al. (2023); Kraemer, Niebel, and Reiberg (2023); Braud, Fromentoux, Radier, and Le Grand (2021)
Matchmaking	Connecting actors, matching demand and supply.	Susha et al. (2023); Kraemer et al. (2023)
Auditing	Federation of the ecosystem, ensuring the policies set are respected, by e.g., providing quality or compliance labels.	Susha et al. (2023); Kraemer et al. (2023); Braud et al. (2021)
Software-as-a-Service	Software delivery on-demand, provisioning of (pluggable) components meant to perform specific data processing aspects or functionalities.	Hilbig, Etsiwah, and Hecht (2018); Sorescu (2017)
Regulation compliance	Unburdening or automation of regulatory compliance.	Kraemer et al. (2023); Braud et al. (2021); Otto, ten Hompel, and Wrobel (2022)
Data standardisation	Enhancing data using metadata, modelling data, and adapting data to standards to ease communication.	Susha et al. (2023); Otto et al. (2022)
Access granting mechanisms	A trusted service that negotiates terms of access to data. Value comes from the legitimacy of the process. A solution for sharing data (including sensitive data) in a secure, targeted, and controlled manner.	Braud et al. (2021); Susha, Flipsen, Agahari, and de Reuver (2020); Schweihoff, Jussen, Stachon, and Möller (2022)
Data-as-a-service	Aggregation service of previously siloed data from multiple sources into a central data repository to ease discoverability of data.	Sorescu (2017); Susha et al. (2020); Hartmann et al. (2016)
Data analytics	Services that generate and derive insights, advice, and visualisations.	Bulger et al. (2014); Sorescu (2017); Susha et al. (2020); Hartmann et al. (2016)
Clearinghouse	A middleman service that approves data exchanges according to policies, e.g., are access rights to data respected.	Susha et al. (2020); Wernick, Olk, and von Grafenstein (2020)
Data broker	Monetising proprietary first-party and personal data. The data is not necessarily altered, therefore it can be a direct retailer of data.	Bulger et al. (2014); Sorescu (2017); Hartmann et al. (2016); Schweihoff et al. (2022)
Governance tools	Tools and expertise that allow ecosystem actors to mitigate potential conflicts of interest, build trust, manage expectations, craft a shared vision, and provide them with coordinating policies.	Susha et al. (2023); Braud et al. (2021); Otto et al. (2022); Susha et al. (2020)

In order to gather a representative sample of the data ecosystem initiator market, several search terms were used to systematically scan this market. This search must represent the current trends in this market, and was therefore not only limited to the initial keyword "Data Ecosystem". The keywords used were adapted to what the first search results delivered as synonyms of data ecosystem, and other common terms in this domain like "Data Spaces". Furthermore, this search must represent an organisation interested in making use of a data ecosystem, who would likely adapt their search to find alternatives. The used terms can be found in the dataset, listed next to each initiator found. The datafile itself is available at de Mildt, Verbrugge, and Colle (2023). Given our research institute is located in Flanders, we cannot ignore Flanders positioning in data ecosystem development, represented by several Flemish/Belgian initiators in the dataset. However, it must be noted the dataset includes many more international examples in order not to impede on its fair representation of the market. In this analysis we determined the current initiator's maturity, the ecosystem scope in which they (or their customers) act, the main service that is offered by the initiator, and characterisation of the ecosystems they set up according to the ones extracted in Section 2.4. Apart from the initiator's websites, the sources used to gather this data were meetings and talks with representatives, technology and business whitepapers, and customer stories mentioning the purposes of the initiators' services for these customers. For each data field not discussed in Section 2.4, the possible values and their meaning (which are the criteria used to categorise them in a certain way) will be explained in the next paragraphs. It must be noted that each data field's value for each initiator can change over time, and thus represents a snapshot of how they were operating at the time of the analysis. To ensure a robust categorisation was made here, several iterations were done to re-evaluate the categorisation and careful consideration of border cases was done. These iterations not only included discussions with, and checks by, the co-authors, but also presenting the work in project meetings (SolidLab, as mentioned in the acknowledgements) where a dozen other academics researching the same data-field were present and able to give feedback. These academics predominantly had a background in economy, sociology and law. This resulted in a wide range of organisations, as illustrated by the distribution of parameters describing the selected data ecosystem initiators in Table 4.

Stage The maturity, called the stage, of the initiator was determined. This data field reflects whether the initiator's operations are still in an early stage or have evolved into a scalable, productised phase. Consequently, this data is time-sensitive, and was based on information available in and before August/September 2023. This field can take on four values being: conceptual stage, pilot cases, non-pilot cases, and licencing/subscription product. In the conceptual stage, there is no use case or product information available

Table 4
Description of the dataset of 100+ organisations that act as data ecosystem initiators.

Data field	Value	Occurrences
Stage	licencing product	57
	pilot cases	24
	non-pilot cases	13
	conceptual stage	9
Scope	Single-organisation	34
	Application domain	28
	Cross-domain	41
Main service(s)	Data-as-a-service	47
	Data analytics	17
	Governance tools	16
	Infrastructure	12
	Matchmaking	11
	Data standardisation	8
	Data broker	7
	Consultancy	4
	Software-as-a-Service	4
	Access granting mechanisms	4
	Auditing	3
	Regulation compliance	2
	Clearinghouse	2

yet, and consequently the initiator mainly pitches a new idea they are working on. When there are a few use cases available, that showcase the services offered, and typically result from a research phase, they are in a pilot case stage. However, if a company moved past these showcases (e.g. operating on a consultancy basis), but has no software product, this is considered as the non-pilot case stage. Finally, once there is a reusable, out-of-the-box software product, which typically has a pre-set pricing model, the organisation is in the licencing/subscription product stage.

Ecosystem scope Three main types of ecosystems are considered here, based on the scope of the ecosystem: (1) Single organisation scope, (2) Application domain specific scope, (3) Cross-domain scope. This scope defines the boundaries of the ecosystem regarding who are participating actors and who are considered outsiders. First, there are the ecosystems connecting a single organisation. Such ecosystems are typically confined within an organisation's walls, sharing data across departments. Commonly, their data sharing solution is based on a proprietary product, relating to the platform-centric ecosystems described by Susha et al. (2023). Secondly, the application domain specific ecosystems target industry verticals within application domains. They promote co-operation outside of organisation boundaries, albeit within specific domains. There is usually a common guideline or standard based on specific industry needs. Finally, the cross-domain ecosystems represent a more divergent set of ecosystems. Such ecosystems relate to the intermediary-based ecosystems of Susha et al. (2023), who facilitate the sharing of data while taking on a neutral role. Ideally, these alleviate the potential of each piece of data and reuse it for multiple, possibly unrelated, use cases. Commonly accepted standards and practices are generally necessary in order for such ecosystems to function.

Main service offering Finally, each of these initiators offer one or more services that set-up and facilitate data ecosystems. Typically, they focus their efforts towards marketing one or two main services, which is represented by this data field. Using this data field, a distinction is thus made between different service providers and which ones offer similar solutions. Since this is a high-level description, it must be noted this does not negate that each initiator still has their own ideas and implementations of these services, and they can still offer additional relevant services. However, the choice of including only their main service(s), rather than an exhaustive list, allows for a clear distinction between initiators in the market, as several offer a multitude of additional services. The values that can be taken on are those shown in Table 3.

3.2. Market data analysis methodology

Having established which information was used to construct the dataset, an explanation is appropriate on how the data was analysed to address the research question put forward. Firstly, several examples that are representative for the dataset are described in more detail. This allows for discussing the reasoning behind setting up these ecosystems, describing the characteristics of the technologies they use, what purposes data serves in their ecosystems, the full set of services offered, and what the arguments are for categorising these ecosystems a certain way. The added value of this detailed discussion is that it renders the dataset more tangible, to be able to generalise by analysing the full dataset.

The study is in essence a broad range of qualitative case studies, categorising organisations and companies based on the often ambiguous and marketing-heavy information available on them (Askarzai & Unhelkar, 2017; Yin, 2018). Similar principles as previous data ecosystem case studies like Susha et al. (2023) on how to interpret such qualitative case studies are considered here. The same applies to general principles of interpretive studies (Klein & Myers, 1999); (Walsham, 1995). However, where these in-depth studies of a limited set (typically only 2 to 4 cases, for example as done by Susha et al. (2023)) are common, the focus here is on a more broad overview of this data ecosystem market. This given the fact, to the best of our knowledge, such a study

has not been performed yet. Therefore, while similar, the methodology employed is less in-depth than those targeted case studies (for example not all of them are based on talking to company representatives, and not all meetings with company representatives are in-depth interviews). Specifically, the goal is not to generalise a certain category from one case. Rather, it is to find the nuance within these categories by studying a multitude of cases to get a more general idea of what this market is like. In this sense it is an inductive, bottom-up approach, as described by Askarzai and Unhelkar (2017) with a wide-angle and deep-angle lens examining the phenomenon of data ecosystems in the market. Several other attributes that (Askarzai & Unhelkar, 2017) outline of qualitative studies are discussed. The behaviour of these data ecosystem initiators can be highly contextual and dynamic, of which the goal is mainly to gain descriptive understanding of what this market looks like. This leads to the data analysis itself to be a thematic, pattern analysis to test the patterns that are described in literature of benefits and characteristics (Section 2.5) to reality in the market. To ensure the rigour and reliability of the study, principles of Yin (2018) to evaluate the strength of the case were applied: construct, internal and external validity, reliability, and rival explanations. Construct validity was considered by using multiple sources of evidence (interviews, websites, talks by company representatives…) where possible, and having multiple review cycles during the research set-up with researchers from varying disciplines. To ensure internal validity explanations were built by combining literature, regulatory documents, and policy vision and goals. This way, multiple perspectives were taken into account to identify patterns. External validity was addressed by applying the established data ecosystem theory to define the concept and its characteristics, rather than building new ones ourselves. Reliability was tackled by publishing the data on which the studied organisations can be found, so that the study can be replicated at a later point in time or built upon, given the evolving nature of these organisations. Finally, in the discussion, rival explanations are addressed. More context in the specifics of this analysis for these attributes and principles is given in the following paragraph.

By a first conversation with company representatives, an introductory presentation on what they do, investigating their website… a first hypothesis is formed on how the company operates. In a second phase, this hypothesis gets tested by delving deeper into their operations. Technical documentation, use cases they have implemented, and the standards and protocols them employ for them should be inspected. In this phase it is crucial to critically approach the available documentation, and where possible meet up with a company representative to validate the hypothesis made. For example, simply because they state their solution is interoperable and can be used to create open ecosystems, does not mean they actually do. At that point, it is checked if their use cases indeed include several diverse stakeholders. Ideally, ones that are not explicitly named: e.g. not a company name, but rather a type of stakeholder like “medical practitioners”. Such pointers give strong indication that it is not known beforehand what actors are or will be present in the ecosystem, and these actors can still enter the ecosystem after it has been established. This way, an objective and robust categorisation can be made, that is not biased in favour towards those companies who market their product in a more positive way than others. Similar indications for improved data quality are whether certain data standards are implemented. Ideally, open ones, or at least transparent communication on which standards are used. A critical look must then be taken at whether these standards are commonly used (within the specific sector of the actor, if relevant). Adding more metadata to data by transforming it into a standard that is not commonly used, is likely to clutter the data (decreasing the quality) for other actors with whom it is shared. Additionally, having an open ecosystem can promote feedback loops since more actors are actively involved in the ecosystem, strengthening the case that irregularities in data become less likely. A fair power distribution being part of their governance strategy, by for example distributing data under the control of different entities, also indicates that data is contributed by various parties. This increases again possible feedback loops, and data quality being improved by consensus. Several of these criteria indicate that data quality improves by making use of the data ecosystem at hand. Regarding innovation, the strategy for data of the Commission provides a relevant indication, given they envision innovation as an end results of the data sharing they promote. Two main indicators are looked for here: data reuse and having a purpose for data. This requires looking into the details of use cases. A first essential criterion is having multiple use cases, simply put, if there is only one use cases there cannot be any data reuse in different use cases. Additionally, there should be given a purpose to the same data over different use cases. Next to this, the services of the ecosystem initiators should not be limited to services that allow actors to share data. While being able to share data (e.g. by offering a data store from which it can be shared) is essential, it is in the purposes for that data after it has been shared where innovation occurs (European Commission, 2024b). Analytical services that make use of shared data are therefore needed for innovation to happen. It must be noted that, due to the dynamic nature of data ecosystems, this study provides a snapshot of the characteristics of each ecosystem, at the time of constructing the dataset. Each of the characteristics can also influence another characteristic. For a more closed ecosystem, it is for example less likely worth it to invest in application independent open standards and protocols. However, the scope of this work focusses on the benefits, rather than exploring the relationship between ecosystem characteristics. To extract additional information from these sources on organisations and companies, a value proposition analysis is done. This value proposition analysis draws inspiration from Value Proposition Design by Osterwalder et al. (2014) in order to extract benefits from the services offered. The main element from this methodology used is the idea of gain creators and pain resolvers, showing which jobs of customers create new opportunities or are made easier by the initiator’s products.

The primary insights from the dataset were inferred by, per ecosystem scope, figuring out what the distribution of each characteristic’s values is. Answering questions like are open ecosystems more common in cross-domain scoped data ecosystems than in single-organisation data ecosystems? This was thus done for the characteristics listed in Section 2.4, the main services offered, and the stages the ecosystems were at, as discussed in Section 3 on the different fields in the dataset. What this reveals is how the ecosystems are typically set up per scope, revealing their similarities and differences. Grouping the different services per ecosystem scope reveals how the value propositions and according benefits differ. Finally, the stages provide information on which ecosystem scopes are operating in scalable fashion, reusing data, and which are not (yet), in order to link back to possible problems with data ecosystems that might not be achieving certain expected benefits.

4. Results

Having established the background knowledge on the expected benefits and characteristics of data ecosystems, and the data set of the relevant organisations active therein, it is now possible to analyse the resulting data set. This data set is analysed both by looking at the general market and several tangible examples in the data set. For the general market analysis, the differences between data ecosystem types are laid out and compared to one another, as well as their envisioned theoretic benefits. Regarding the examples in the data set that are discussed in more detail, specific reasoning and context is added to these data ecosystem players to reveal case-specific insights.

4.1. Market analysis

The market analysis performed results in several insights into the dataset of market players. Figs. 1, 2, and 3 provide the characteristic insights of these market players in the dataset. These three figures show the occurrence of each characteristic value (Openness: closed or open, Data access: siloed or application independent, and Initiator goal: privacy protection or data sharing of Section 2.4) per ecosystem scope, in the dataset. Fig. 4 show the different services observed in the complete dataset. To derive the necessary conclusions, two trains of thought are used. On one hand, data sharing tools must allow for more open and application independent operations (linking to the characteristics' values). Therefore, if these are commonly offered for a certain ecosystem scope, to a diverse group of actors (e.g. not just to one company per use case), the benefits of more available data, of improved quality, are considered to be observed here. Regarding the innovation benefit, there must be tools offered that allow data analysis, and use cases in which data is reused by multiple actors. Finally, Fig. 5 shows the different maturity stages each ecosystem type is in. Where the single organisation market is dominated by licensed products, there is a shift towards more conceptual and pilot case stages for the broader ecosystem scopes, indicating a lack of use cases in which data is reused from other use cases. Table 5 provides an overview of matches, and mismatches between expected benefits, and actual perceived benefits in the market. The single organisation scope is the only type in which the initiators offer services that increase innovation in practice. However, application domain specific and cross-domain ecosystems carry innovation potential, as there is a consistent observation of open and application independent ecosystems with supporting services therein. The following paragraphs briefly discuss this in more detail, per ecosystem scope.

Single organisation scope. Regarding the single organisation scope, the following characteristics were consistently observed (shown as the first bar in Figs. 1, 2, and 3): (1) Openness: closed, (2) Data access: siloed and (3) Initiator goal: data sharing. The only benefit these should offer is thus increased innovation due to data sharing if we link back to the theoretic benefits from Section 2.5. When looking at the perceived services in the single organisation market, these consistently contained data analytics. Specifically, focusing on making use of newly combinable data to bring forth novel insights, which leads to a benefit of increased innovation. Sargent and Denniston (2023) strengthen the claim that data analytics should lead to innovation by stating that data science assets show the most growth in the information value chain. Secondly, due to there now being a single-source-of-truth among departments, the quality of the data also improves as a perceived benefit.

Application domain scope. The application domain held the following characteristics (shown as the second bar in Figs. 1, 2, and 3): (1) Openness: open, (2) Data access: siloed/application independence and (3) Initiator goal: data sharing/privacy protection. Different ecosystem types were perceived here, leading to all three benefits of improved quality, more available data and increased innovation, to be expected in theory. When checking the services perceived in the market, these ecosystems often offer improved quality by auditing the ecosystem's actors and making sure policies are respected. Next to this, they typically increase the amount of available data by offering (support for) aggregation services. However, when it comes to increased innovation, this is left to the initiator's customers themselves who then need to build, for example, relevant data analytic services on top of the base aggregation and governance services.

Cross-domain scope. Finally, when looking at the cross-domain scope there is a clear consensus on innovators building these ecosystems (shown as the third bar in Figs. 1, 2, and 3): (1) Openness: open, (2) Data access: application independence and (3) Initiator goal: data sharing. Therefore, again, all three benefits should be expected theoretically. Again, looking at the practical implementation's in the market, the perceived focus is on supporting data services. These include standardisation, access granting mechanisms, data transaction support and data aggregation. This leads to both improved quality (e.g. standardisation and transaction support) and more available data (e.g. access granting mechanisms and data aggregation). Again, the innovation is left to the users of the data, rather than the initiator's typically offering, for example, analytical insights. However, contrary to the application domain specific scope, there typically is no clear supporting entity. Orchestration of the ecosystem is more of a distributed process of the ecosystem actors, with a facilitating role for the data intermediaries in this process.

The single organisation scope is where initiators consistently offer innovative services. While theory expects all ecosystem approaches to innovate by acting on data, only the single organisation scope offers services that gather new insights from data by, for example, co-producing between departments and reducing operational costs. Both the application domain specific and cross-domain ecosystem initiators leave this up to the ecosystem actors, focusing on, among others, governing services. Fig. 4 supports this statement, with one of the dominating services being data analytics in the single organisation scope, while governance tools are more common among the other two ecosystem scopes. A plausible explanation is the fact that the single-organisation experience less trouble accessing data, due to being closed and siloed, as shown in Figs. 1 and 2. While this limits the potential, it provides a

Table 5
Results of challenging theoretical benefits: Benefits observed (✓) and not observed (×) in theory and practice (respectively). Two similar symbols therefore denote a match.

	Improved quality	More available data	Increased innovation
Single organisation	×✓	××	✓✓
Application domain specific	✓✓	✓✓	✓×
Cross-domain	✓✓	✓✓	✓×

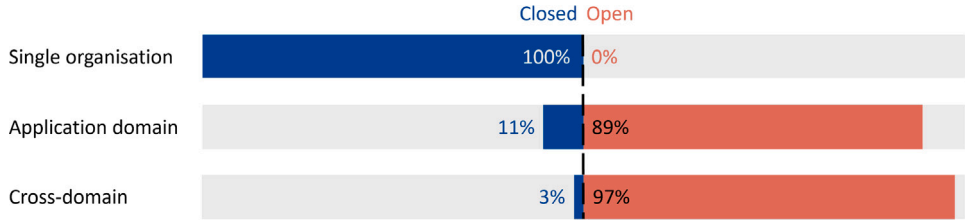

Fig. 1. Observed Openness from the characteristics of Section 2.4 in the market of data ecosystem initiators, shown for each ecosystem scope.

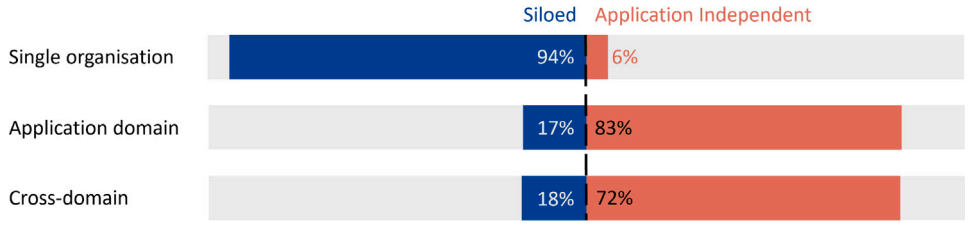

Fig. 2. Observed Data access from the characteristics of Section 2.4 in the market of data ecosystem initiators, shown for each ecosystem scope.

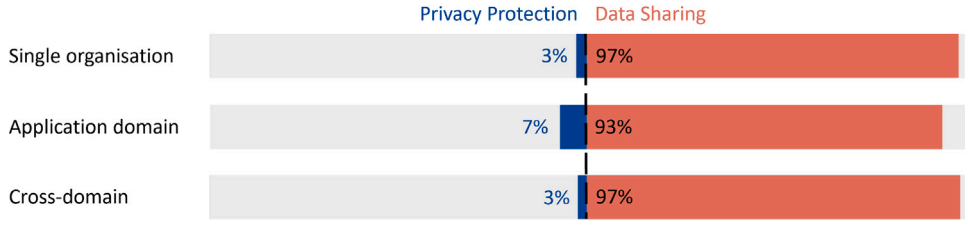

Fig. 3. Observed Initiator goal from the characteristics of Section 2.4 in the market of data ecosystem initiators, shown for each ecosystem scope.

pragmatic way of exploiting data in a limited ecosystem.

Application domain specific and cross-domain ecosystems carry innovation potential We can observe a clear dominance for open, application independent ecosystems for both ecosystem scopes in Figs. 1 and 2. Both ecosystem types offer services like data-as-a-service, governance tools, and data standardisation, as shown in Fig. 4. These services adhere to the benefits of making more data, of higher quality, available. Therefore, (additional) data both becomes accessible and reusable for the actors in the ecosystem, compared to before adopting an ecosystem approach. When examining the stages these ecosystems are typically at, shown by Fig. 5, pilot cases and the conceptual stage appear more often. This reveals that steps are to be taken in order to provide scalable solutions, let alone achieve this innovation potential. Referring back to the application domain and cross-domain ecosystems in the dataset and Section 4.2, Gaia-X, Athumi, Public Spaces, and the Flemish Smart Data Space are (partially) backed by public funding. This shows clear belief in their potential, however, this does raise the concern of funding not meeting its intended innovation goal (yet).

4.2. Data ecosystem market examples

As mentioned in the methodology of Section 3.2, in order to reveal how ecosystems offer benefits, they must be tested against the theoretic expectations of Section 2.5. To make the link between theoretical expectations and practical market implementations more tangible, several market representative examples from the dataset are covered in more detail. The purpose is to better understand the dataset, and the general insight resulting from it. These examples are structured per ecosystem scope, in order to make an easy comparison between these ecosystem types. Their characterisation, based on Section 2.4, and an overview of the services they offer is given, based on Section 2.6. Finally, the value offered by their services is discussed.

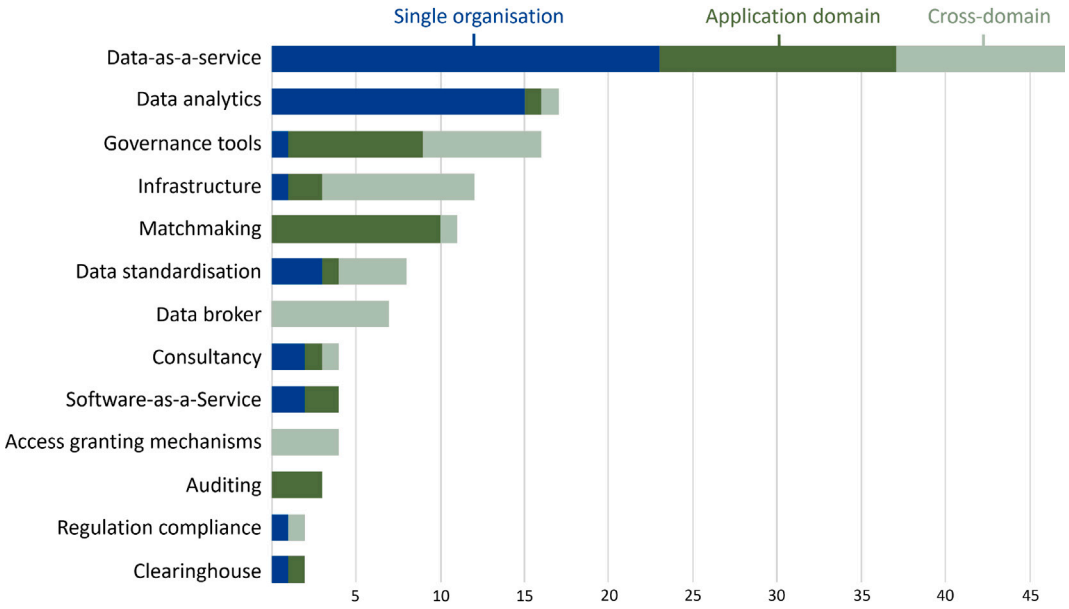

Fig. 4. Observed main service offerings by data ecosystem initiators, shown for each ecosystem scope. A description of each service is found in Table 3.

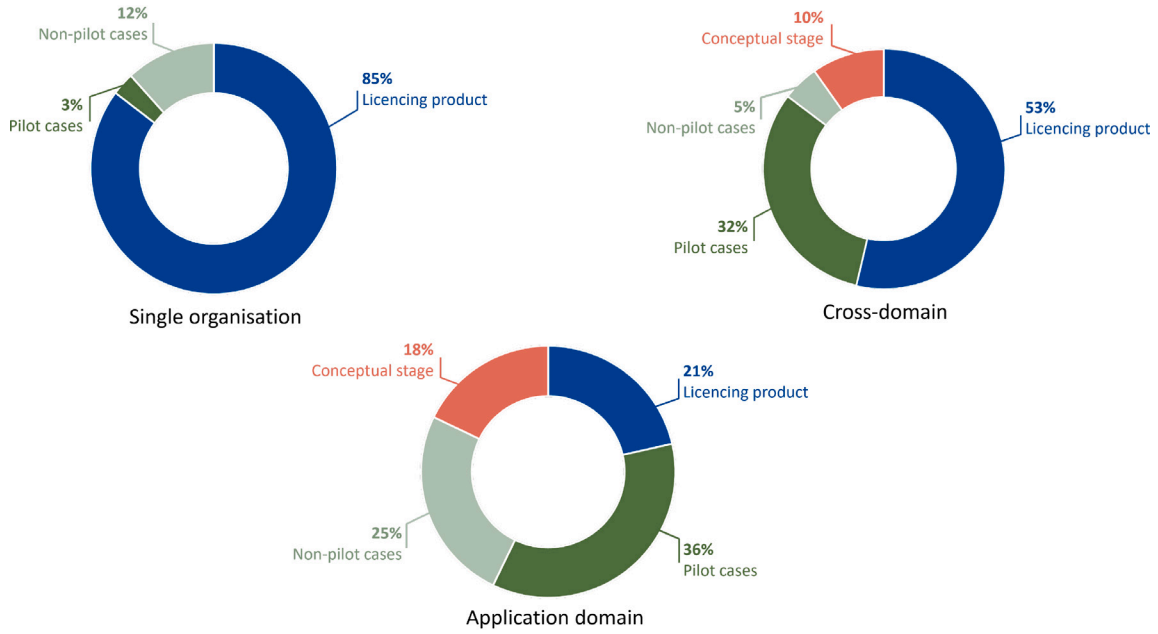

Fig. 5. Observed stages data ecosystem initiators' products are in, shown per ecosystem scope. A description of each stage is found in Section 3.

4.2.1. Single organisation scope

Snowflake. Snowflake (*Snowflake*, 2023) as a platform provider, facilitates data sharing for organisations using a variety of tools. Providing infrastructure, standardisation, central data points, analytical services and governance tools, to name the most prominent ones. Typically, connecting data departments with departments like marketing. As a platform, Snowflake does not always partake actively in the ecosystem, but often operates as a (white-label) technology provider. In their typical customer cases, ecosystems are closed, the data is still stored in silos and only a Snowflake product connects these silos. However, Snowflake looks for clients that want to share data, albeit within limited organisational or partnership boundaries. Focusing their marketing strategy on licensing products serving multiple industry domains, there is a some surface level degree of transparency on how data is combined and reused for varying purposes. Reflecting upon their data ecosystem offering, and how this fits into policy regulation's goals of innovation in a single market for data, there is a mismatch here. On one hand, there is an abundance of use cases, in which the focus is on

Table 6
Analysed examples and their services offered in the single organisation scope.

Example	Openness	Data access	Initiator goal	Services
Snowflake	Closed	Siloed	Data sharing	Consultancy - Infrastructure - SaaS - Data standardisation - Data-as-a-service - Data analytics - Governance tools
HubSpot	Closed	Siloed	Data sharing	SaaS - Data-as-a-service - Data analytics
Census	Closed	Siloed	Data sharing	SaaS - Data-as-a-service - Data analytics

reuse of data within companies. There is also no lack of innovative industries that make use of their products, improving healthcare being just one example where Snowflake is active (Snowflake, 2024), that fits into the EU's strategy. However, while the innovation is there, the openness and actual sharing of data in a wider scoped ecosystem is lacking. Data sharing and usage is dependent on Snowflake's proprietary products (=data silos). Given this lack of interoperable data sharing possibilities, they do not fully contribute to the single market strategy.

HubSpot. HubSpot (2023) focusses on providing a platform to manage customer data and automate the business processes across a company. Some services they offer are central data points, marketing insights and analytics, and data management tools. Targetting companies internal processes, and connecting silos therein (often simply syncing data stores rather than using a new central point), these are typically closed ecosystems. Sharing data within their organisation, and focusing on existing processes and needs a company has when it comes to data management first. Offering licensed products, that focus on process optimisation and novel insights (e.g. Marketing software for generating leads), also for employees with little IT-knowledge, they make clear from the start for which purposes data will be used. When reflecting upon HubSpot's offering to set up data ecosystems, an abundance of use cases is present (HubSpot, 2024). If the marketing figures on their results are to be believed, data is (re)used here both to innovate with it, and save costs to make room for more innovative purposes to allocate funds to. However, each case is represented as "Company X's use case" "Company Y's use case" making it clear that data sharing is again limited to company boundaries. Therefore, not contributing to a single market, given dependency on HubSpot's proprietary products, limits data sharing outside their data ecosystems' scope.

Census. A similar offering is given by Census (*Census*, 2023), which allows businesses to integrate their data warehouses, as the Census website puts it, with marketing tools. This includes integration services for a variety of common business and marketing tools, as well as analytics and management facilities. As their offering is similar to the other two examples named, they too focus on organisational ecosystems. They also set up closed ecosystems, essentially connecting platforms, being an intermediary platform themselves. Next to this, they build mappers that do not fundamentally change the paradigm on how data is stored in silos, but connects them one-on-one. Finally, the goal of census is to make their customers able to share data they previously had a hard time transferring between departments. Given the similarity of their offering to the other two companies discussed in this section, a similar reflection can be made. They focus on data usability, but lack data sharing tools that go beyond their own ecosystem scope.

Value proposition analysis. Given the similarities to these services and ecosystem characteristics, these share a similar value proposition analysis. The most prominent gain creators are the analytical insights that can be extracted when these services make data available to different teams. The full dataset confirms this argument. Fig. 4 shows a dominance of data-as-a-service and data analytic services, which carries this combination of making data available and extracting useable insights from that data. Accordingly, data becomes reusable across an organisation and co-production of value using knowledge from different teams is made possible. Optimisation of processes, by making sure there is a common data source that is reusable, is typically one of their main pain relievers. One such an example of a gain creator they offer is automated discovery of insights, or better said tools that allow this. Summarised, regarding the benefits named in the theoretic analysis: no additional data becomes available (it was not useable, but the data did not change organisation), but additional data becomes useable due to a quality increase (e.g. single-source-of-truth) and novel insights are extracted, meaning innovative services are offered. Single organisation scope ecosystems are typically closed, siloed ecosystems aimed at data sharing. In line with the larger dataset, data analytics and data-as-a-service are offered in all three of them (see Table 6).

4.2.2. Application domain scope

Gaia-x. Gaia-X (2023) is an international, European Commision backed, initiative that aims to bring together different European member states to increase the adoption of cloud services and data exchanges by European businesses. As an organisation, they support the development of what they call data spaces, essentially industry verticals for data sharing. Developing standards and guidelines, Gaia-X allows for service providers to build platforms based on data sharing. However, Gaia-X does not offer the platform implementation itself. That is left to the service providers who want to build such a platform, according to common standards. One example is that of Gaia-X's federation (governance tool) service and catalogue. Gaia-X aims to set up open ecosystems in which actors can participate and, through the federation services, receive e.g. trusted labels like GDPR compliance. However, they do not interfere with the data storage, remaining neutral on data access. As with the other ecosystems analysed until now, the goal is still to create additional economic value by sharing data. Essentially, the customer's of Gaia-X are platform providers, who in their turn still typically operate as intermediaries, like Prometheus-X or Catena-X (*Gaia-X lighthouse projects*, 2023). Reflecting on their products it is not surprising, given this initiative is backed by the Commission, they fit into the single market strategy. Through

their catalogue and accompanying federation services, they focus on data findability and query-ability. Therefore, they offer the tools for open and application independent sharing of data, across sector boundaries and with any actor. However, looking at the platforms that partake in the Gaia-X ecosystem, clear innovative results remain to be seen. Many of them are still in development or in pilot phases. In future development, breaking the sectorial boundaries could be one way to increase actual data reuse. While the tools are there to reuse data, the operational use cases still lack innovative data reuse across use cases for now.

HR-domain MVP by athumi. One technology that allows data ecosystems to be developed is that of Solid (Sambra et al., 2016). Solid technology allows development of data ecosystems as a set of open web standards that promote interoperability, in combination with personal data storage technology. *Athumi* (2023) is one of the companies, recently instated, that makes use of this technology for some of its ecosystems it sets up. Essentially supporting data sharing ecosystems, they set up several domain specific ecosystems. One of them is the HR-domain, in which they play a keystone role as a connector between private parties and government databases for sharing diplomas. Some other services they provide are infrastructure, auditing the ecosystem and standardisation services. Involving citizens, through personal data stores, in their data sharing initiatives, they create an open ecosystem. They offer their services both to platforms, and in cooperation with (government) platforms, meaning within the Athumi set-up ecosystems there are different possible roles for platforms. One example in their diploma-case is the MijnBurgerProfiel platform being an intermediary that allows citizens to grant and manage data access permissions to job-matching platforms. Given they make use of Solid, the essential idea is to use personal data stores for each citizen. This means they fundamentally break open silos and redistribute this data across the ecosystems they build. Next to this, Athumi embraces a Linked Data standard, OSLO (*OSLO standards*, 2023), which results in standards that are reusable after pilot use cases have been implemented. Therefore, Athumi tries to jumpstart new economies by facilitating data sharing. As the positioning of Flanders was mentioned in the data gathering, this is one such representative case for Flanders. Given its presence in the heart of the EU, Flanders is subject to the same regulations, and economic focus on data (sharing). To name one example, Flemish public and private organisations, local government... have focused on publishing open data in the latest years (Digitaal Vlaanderen, 2024a). Next to this, the government openly emphasises the importance of privacy concerns and the economic potential of fair handling of data (Digitaal Vlaanderen, 2024b). This is one such case of Flanders showing a push from government towards open data sharing initiatives that focus on application independence. In the general data ecosystem landscape, Flanders thus can be positioned as setting up data sharing in more open fashion across applications. However, their current number of use cases is limited, and consequently so is the data reuse in their ecosystem. In regard to the single market policy goal, Athumi sets up an ecosystem that has all the necessary tools to contribute to such a market. However, this focus on the intermediary step, results in them still needing to tackle data reuse adoption barriers. True innovation is still lacking in their current phase of development, and a focus on data reuse should be a crucial target in further steps. This case also highlights the issues still present with geographical differences in data sharing. Initially, specific regions are targeted with data sharing use cases, like Flanders. Since Flanders does not even cover the whole country of Belgium, it shows how significant effort is required before expanding across language barriers, government jurisdictions and international borders, even within the EU. The legal implications of crossing such borders seems the major hurdle to take, rather than technical issues. In this specific HR-domain case, local Belgian and Flemish law requires possibly different data than other EU member states when applying for a job, and registering a valid contract upon recruitment. Next to this, even within a specific region, international applicants, often times have a different identification system than a Belgian eID card. Even for other member state citizens, this often times requires additional mapping and verifications of identity to share diploma data in a verified manner.

Public spaces. In the Netherlands, the Public Spaces (*Public Spaces*, 2023) initiative was created based on public values, and is therefore a bit atypical from the earlier mentioned organisations. The services Public Spaces offer focus on organising collaborative thinking to support development and research of platforms that support distributed governance. One example supported by Public Values is that of PubHubs (van Dijck & Jacobs, 2024). However, being a bit of an odd-one-out, the application domain is not an industry vertical but rather a "public" vertical. What must be seen as a public vertical is for example being part of a Hub of a news channel, museums or one of your municipality. The Public Values initiative offers support to anyone wanting to create ecosystems based on public values, rather than create new economic possibilities. Therefore, the ecosystems they support are focussed on privacy protection rather than sharing data for economic gains. A main challenge here is that this protective stance might hinder data reuse and combining data from different sources, as outsiders might not know which data is reusable nor do they know for which purposes it can be used. As public values are central here, the ecosystems they facilitate are open and based on breaking down silos and redistributing data across the ecosystem. When reflecting upon their realisations, the PubHubs platform is still in early stages (PubHubs, 2024). It remains a question if innovation will be possible if the main goal is to protect their citizens. While a noble goal, and it certainly has its role in a single data market, future development will have to tell how such a platform can become a sustainable solution.

Value proposition analysis. Analysing these services the focus clearly shifts here, towards the gain creator of being able to collaborate with industry partners more efficiently. Adopting common standards and, what is considered good practices are recurring elements for each of the mentioned initiators. The initiators often being keystone actors that audit the ecosystem, relieving the pain of making sure partners are not breaking policies. Knowing these keystones are often either closely collaborating with, or part of, government organisations provides the pain reliever of being compliant with (latest) regulations. This way, the pain of integrating with specific products with an added cost when scaling to new partners or reusing data for new purposes as a consequence, rather than standards and technology, should be relieved at least on an application domain level. Linking back to the full dataset, Fig. 4 reflects this in the variety of the aforementioned services offered. Summarised, regarding the benefits named in the theoretic analysis: additional

Table 7
Analysed examples and their services offered in the application domain specific scope.

Example	Openness	Data Access	Initiator Goal	Services
Gaia-X	Open	Both	Data sharing	Matchmaking - Auditing - Data-as-a-service - Governance tools
HR-domain MVP	Open	Application independence	Data sharing	Infrastructure - Auditing - Data standardisation - Access granting mechanisms - Data-as-a-service - Governance tools
Public Spaces	Open	Application independence	Privacy protection	Auditing - Governance tools

data becomes available that is now being shared among industry partners and collaborations, the ecosystem is audited meaning the quality of data increases, but innovation is only indirectly stimulated. New connections are made possible, but the actual innovative services are left to the actors in the ecosystem themselves, typically platforms having roles as intermediaries themselves or service providers. Application domain ecosystems are a nuanced ecosystem type having open, but both siloed and application independent ecosystems aimed at both data sharing and privacy protection. In line with the larger dataset, data-as-a-service, governance tools, and auditing ecosystems are among the dominant services offered (see Table 7).

4.2.3. Cross-domain scope

DataVaults. The *DataVaults.eu* (2023) initiative aims to create an ecosystem of personal data sharing. Several of their offerings include data management, data brokerage, and data anonymisation services. Putting the end-users at the centre of their ecosystems, they are open for anyone to participate in. End-users must be given back control, according to the DataVaults initiative, meaning traditional silos are not part of this ecosystem. Finally, while preserving privacy is important, the initiator goal is still to share data in order to create future monetisation streams. A set of demonstrators is listed by DataVaults, from which it is clear different types of customers can make use of the DataVaults platform. Some of these customers are intermediary platforms themselves, as well as end-service providers. Apart from these demonstrators, it is unclear which other projects make use of DataVaults technology, if any. One cause might be that there is limited information on the data being used in these demonstrators (e.g. which format, how to access it...) for future service providers who would be interested. Using these demonstrator cases to make a reflection on if they achieve the goals of innovation in a single European data market, they do not (yet) move past these demonstrators (DataVaults Project, 2024). In each of the cases, there is a mention of sharing data with multiple relevant stakeholders (and thus reusing it), moving past the stage of only serving the initial partner mentioned in the demonstrators. However, there is no indication they actually move past this stage. Therefore, the tools and willingness to fit into a single market are there, but the innovative goals are lacking.

Meeco. *Meeco* (2023) also offer a personal data solution in their vaults. Combining this with identity management, a data sharing platform, consent management tools and verifiable credential services, creating a personal data ecosystem. Again, openness is central in such an ecosystem, where authentication and verification is provided through identity providers and VCs. Aiming to give back data control to the end-user, silos are broken down. Lastly, the initiator goal of Meeco towards their customers is to stimulate data sharing to create new economic possibilities. A representative use case for Meeco's vaults, listed on their website, is their integration into the KBC-banking platform, in which end-users of KBC can upload and safely share documents and data of their choice. Therefore, the typical customers of Meeco are platforms and intermediary services. However, the endless potential of data that can be stored in their vaults also lacks a clear description of which data can be shared and reused for interested parties. Reflecting on their offering, they build on open standards, some examples are OpenID (OpenID Foundation, 2024) and EBSI (European Commission, 2024a), of which EBSI is a Commission initiative. Therefore, it is no surprise they fit into their strategy. When looking at their use cases, while still a limited set, they focus on maximising data reuse (Meeco, 2024). It is unclear if they can bring forth strong arguments for innovative data reuse to date, however the ecosystem data sharing tools are there.

iGrant. *iGrant.io* (2023) provides again a similar offering, using Solid technology to store personal data, they also seek to give back control to end-users. They focus on regulatory compliance, infrastructure in the form of Solid pods and verifiable credential services. Similar characteristics can be granted to this ecosystem as for the other two personal data vault initiatives (DataVaults and Meeco) mentioned before. Looking into use cases for iGrant 's vaults, most of them are showcases in the form of a proof-of-concept without actual users involved. However, the information listed shows reuse of data in different showcases and using different intermediaries for additional services. Focusing on explaining which data is being reused or even combined from their different showcases, there is a certain degree of transparency towards both end-users and service providers on the potential of their solution. Given their focus on tools for data reuse, they fit the single market strategy in the sense of open and application independent sharing of data. When looking into their use cases, there are multiple that focus on sharing some accredited personal data like a passport, a MyData profile or an EBSI social security pass (MyData Global, 2024); European Commission (2024a); iGrant.io (2024). However, results or evidence is lacking on whether these use cases have actual data consumers. Again, a similar reflection can be made that they fit the strategy in terms of having the right tools for creating a single data market, but do not yet clearly reach the innovation potential.

Flemish smart data space. Finally, the Flemish Smart Data Space (VSDS in Dutch) (*Flemish Smart Data Space*, 2023) represents another cross-domain ecosystem initiator. They focus on organisations and companies who want to share dynamic data (e.g. sensor-data). Supporting services to make this data shareable include setting governance rules and auditing them, open source building blocks in the form of SaaS services, aggregating data, and supporting standards like Linked Data Event Streams (Lancker et al., 2021). Several

Table 8
Analysed examples and their services offered in the cross-domain scope.

Example	Openness	Data access	Initiator goal	Services
DataVaults	Open	Application independence	Data sharing	Data standardisation - Access granting mechanisms - Data-as-a-service - Data broker - Governance tools
Meeco	Open	Application independence	Data sharing	Infrastructure - Access granting mechanisms - Data-as-a-service - Clearinghouse - Data broker - Governance tools
iGrant	Open	Application independence	Data sharing	Infrastructure - Regulation compliance - Data standardisation - Access granting mechanisms - Data-as-a-service - Governance tools
Flemish Smart Data Space	Open	Application independence	Data sharing	Infrastructure - Matchmaking - Auditing - SaaS - Data Standardisation - Data-as-a-service

(government) platforms use data-input and services of the VSDS to get access to the data they need in interoperable fashion. Next to this, many intermediary actors operate in the VSDS, of which several are platforms themselves. In order to promote the reuse of data, there is a platform build specifically to browse the data, with all required details like formatting, made available through the VSDS. This case emphasises Flanders role in the data sharing landscape. Similar to the HR-domain case in Section 4.2.2, the Flemish Smart Data Space shows the government push in Flanders towards open, application independent data sharing. Specifically, this case shows the focus on building Data-as-a-service to make data available, findable and query-able. The Flemish government does so by building APIs and standardising data, typically accompanied by a platform through which different (government) data sources are connected. However, when looking into the reach that the use cases of the Flemish Smart Data Space have when it comes to data (re)use, this remains limited (Vlaamse overheid, 2024). The Flemish Smart Data Space consists of 6 different, operational, use cases. While this shows certain maturity going past a pilot phase, these use cases currently focus on government to government data sharing. While there are a few exceptions and private companies mentioned, this leaves out many private actors with whom they could share data. One example is the mobility use case, where traffic measurements from central institutes like the Roads and Traffic Agency, statistical offices and local governments are standardised and connected to one another. They hint towards new possibilities and applications (where the actual innovation should occur), in which this data is then actually used, but none such applications are given. Consequently, the VSDS serves as an example of clear intention in Flanders to reach policy goals by setting up data sharing ecosystems, but true innovation remains to be achieved.

Value proposition analysis. If these services are analysed, the benefit continuously put forward is that of being able to share data with endless potential. Making sure this data can be shared by anonymising it, providing the possibility to create verification of it or modelling it in useable formats and providing the infrastructure needed to share and store data are the pain relievers and gain creators here. Again, Fig. 4 shows that several of the common cross-domain ecosystem services mentioned here are present in the larger dataset. Regarding the benefits named in the theoretic analysis: additional data becomes available given it can be shared without clear limitations, the ecosystem is focussed on actors that provide services that make data shareable and usable. This means the quality of data increases, but innovation is only indirectly stimulated, and might even be hindered by a lack of successful use cases resulting in reusable data. Given the complexity of this data sharing, without clear limitations, this proves to be a challenging problem to tackle. For most of the initiatives in this scope, only limited use cases that are operational can be found, which indicates this complexity still hinders innovation. Next to this, transparent communication on what data is available, or in which format, in order to reuse it for future cases is often lacking. Cross-domain ecosystems are typically open, application independent ecosystems aimed at data sharing. In line with the larger dataset, data-as-a-service, data standardisation, and access granting mechanisms are among the most common services offered (see Table 8).

5. Discussion

The resulting market analysis in this study revealed that the hypothesis of data ecosystems having difficulties sharing data that result in innovative applications and large scale data sharing was confirmed. This shows issues with data ecosystems as a managerial reaction by organisations to data (governance) policies (in the EU). Therefore, a discussion of these results both in the context of data ecosystem theory (Section 5.1) and in the light of policy and managerial practicalities 5.2 is appropriate.

5.1. Theoretical implications

This study has several implications that should be taken into consideration in future studies on data ecosystems. Careful consideration is needed before stating the benefits of data ecosystems when performing a study. The common assumption is that employing a data ecosystem will lead to innovation, just because certain technologies and frameworks are used. Verifying if these benefits are perceived should not be overlooked, as is often the case now. Most data ecosystem literature focusses on *how* data ecosystems will bring these benefits, rather than *if* they bring these benefits, before tackling the *how*. This seems to partially be due to a lack of empirical evidence. Given the limited operational data ecosystems that have multiple actors, there is typically only a notion of expected benefits that can be researched. There is a need for clear KPIs that measure innovation and data reuse of data ecosystems. Future research should be conducted that can provide a theoretical foundation of appropriate KPIs for data ecosystem

success. However, the multitude of perspectives of different actors in such ecosystems, does pose a challenge to define the success of a data ecosystem.

More data being available could be such a measure of success. However, there should still be a critical look on if there actually is more data being shared now, than if there would be a more traditional bilateral agreement to share data between organisations. Simply more data being shared, because there was no sharing happening before the ecosystem was set up, could give a possibly false indication of success. Measuring improved quality of data is somewhat subjective in nature, as quality data can have different meanings for different organisations. Typically, these studies fulfil the quality requisite by transforming data into standards. Again, a more critical look is often lacking whether this standard is widely adopted already or has the potential to be widely adopted. If not, the data could be considered more cluttered than before, as an additional layer of formatting adds unwanted complexity.

Additionally, the results show there are different types of data ecosystems, each with their own characteristics. When performing case studies, or any data ecosystem study, this should be taken into account. Consider what characteristics fit the studied ecosystems and reflect whether the benefits that are theoretically assigned to them, are actually perceived ones of these ecosystems. Such critical reflection is often missing in data ecosystem literature. The sole focus is often on the positive aspects, not highlighting negative side effects, like possible higher costs, increased complexity in data task, more parties involved that each need to generate revenue one way or another... which could be reasons why there are such limited number of data ecosystems that have multiple actors, and actually reuse data. Future research must thus be conducted on why these data ecosystems often do not actually get to a stage of scalable data reuse (yet).

5.2. Policy and managerial implications

Data ecosystems require careful consideration when put forward as a solution to the problem of data silos, or as a means to achieve new innovative possibilities. It is often used as a buzzword and linked with removing barriers and bottlenecks for data use and reuse. However, they require much more than just setting up technology (the main focus of the initiators analysed in this work). There seems to be general awareness of additional (non-technical) requirements, such as standardisation and setting up governance frameworks (deciding on whom to share data with, for example). However, an all too positive picture is painted on the economic and societal potential of these ecosystems, since they are mostly based on expected benefits for which there is little empirical evidence. When it comes to the current data policies in Europe, a similar positive picture is painted on the expected outcomes regarding a single market for data, leading to innovation and other benefits. However, the policies they name in this strategy to create a single market for data seem restrictive in nature (e.g. the DGA restricting data intermediaries). These restrictions are a possible cause for limiting possibilities rather than initiating new innovative applications. Future research is needed to analyse these policies in light of actual achieved benefits of data ecosystems, and if they in turn help develop their envisioned single market for data. They seem to result in restricting data reuse in practice (Shabani, 2021), while their goal is said to increase the reuse of data. This contradiction leads to possibly ineffective policies, resulting in ecosystems where reuse of data is lacking. An important oversight therefore seems to be a measurable, long-term strategy for data reuse.

In such a strategy, the role of platforms also needs to be assessed in the context of data ecosystems. The different ecosystems and their services reveal that platforms do not have a single role or use in ecosystems. Fig. 5 shows that many ecosystem solutions consist of reusable software components by the licensing product and non-pilot case stages. Typically, the observed organisations in these stages offer their software components on (white-label) platforms. Fig. 4 provides additional insight into how different services are offered in such platforms. The single-organisation scope initiators typically offer data-as-a-service and analytic platforms. While the application domain focus more on advisory/consultancy/guidance services, of which the customer segments are typically platforms. Finally, in the cross-domain, several common services are meant as backend services for platforms, like infrastructure and access granting mechanisms. Platforms are thus transitioning from not only being a central actor, but also intermediary services in many forms. Consequently, for data ecosystems to persist, platforms must be positioned within them as such. Specifically, given the DGA (*The Data Governance Act*, 2023) and DMA (*The Digital Markets Act*, 2023) are just two recent examples of policies, that are relevant for digital platforms that embody such intermediary roles. At the same time, platforms are one of the main actors that often do provide innovative technology, and need data to be able to offer innovate services, even if they need to remain neutral intermediaries. Ongoing research by Verstraete et al. (2023) supports this argument by showing that different interpretations of the DGA can lead to different business models being affected negatively in the market, depending on whether they fall within the DGA's scope or not.

Coming back to the data reuse strategy, an assessment must also be made of the demand there is for data, and data reuse in the envisioned ecosystem. A data ecosystem initiator should not just solve one data inefficiency for a specific actor, and expect more actors that want/can use the same data for different purposes to appear by itself. One of the reasons this reuse is lacking might also be that currently the reporting on data in existing data ecosystems is not ideal. For many data ecosystems, there is limited or only superficial info available on what is already being done with which data, and what data is available already in the ecosystem. There is a need for transparent information on what your ecosystem is doing, what data is available, who delivers that data, if there are any costs to using it, what format it is in, what the formal requirements are to enter the ecosystem and so on. This provides an opportunity for policymakers to incorporate in policy that such info should be made available, in transparent fashion. Next to the demand that has to be there, the willingness of all actors involved, and potential future actors, to change their ways of operating must be assessed. Why would they be prepared to incur costs and differ from their current strategy on handling data? A general cost–benefit analysis for possible actor types like technology providers and data consumers could provide a first step. Not only in positive situations and under assumptions that there will be many data reuses, making costs possible shared over cases, but also in

pilot situations (where there might be high costs for the one case there is) and worst case scenarios. Such results will give actors who are exploring the idea of setting up or entering a data ecosystem a way to assess what is in it for them. The market insights of Section 4.1 show that efforts to make data reusable are happening, but suffer exactly this limited reuse, given a lack of innovation in the broader ecosystem scopes. Cross-domain ecosystems, which harness the most potential, for example, are largely represented in the market by services like governance tools, infrastructure and access granting mechanisms, as shown by Fig. 4. Consequently, these ecosystems offer many different technologies to make data reusable, but tend to remain agnostic on reusing that data afterwards for additional purposes, often in order to take a neutral stance as an intermediary. Fig. 5 supports this limited reuse of data as both the cross-domain and application domain specific ecosystems, conceptual and pilot case stages combined make up about half of the observed ecosystem solutions. This indicates that the ecosystems set up do not (yet) thrive with multiple purposes for this reusable data. Several ways to improve this could be to communicate openly which type of data is available, how data is formatted, which metadata is needed, or even a new standard that was developed as a consequence. Next to this, trust requirements that would allow access to the data, the actors that provide data, or a governance framework implementation could minimise the difficulties actors and ecosystems face when assessing if a data ecosystem fits their innovative ideas. In the end, actual value is not created in the data ecosystems services, but in the use cases they make possible.

6. Conclusion

In this research, an exploration of different types of data ecosystems, and initiator's services therein, was performed. This was linked to a narrative found in literature that open, application independent ecosystems aimed at data sharing must deliver more data, of higher quality, and lead to innovative services. In this analysis, the hypothesis that the way ecosystems are offered in the market influences the benefits, was confirmed. Essentially, this paper showed that ecosystem's characteristics and scope influence the benefits of an ecosystem approach. Just because data is available and shared, does not mean innovation based on such data will automatically follow.

The benefits attributed to data ecosystems are not all reflected in the market yet. While steps are being taken in the right direction, the potential for innovation is yet to be realised in many ecosystems. Typically, initiators set up ecosystems that deliver tools for making data shareable, but often remain agnostic towards putting that data to use for specific purposes. This neutral take on the ecosystem approach, is commonly found in the more broad ecosystems that carry, but do not yet realise, the most potential. At the same time, it is exactly these more broadly scoped ecosystems that are typically backed by public funding. Questions must be asked if this push, being a result of both this public backing and regulatory policies mentioned in the introduction, needs a correction of some sorts. Current focus on having data shareable and findable rather than focusing on data actually being reused makes it so that many initiatives might not reach mature stages. At the same time, the foreseen economic and social innovation goals are yet to be achieved. One of the crucial next steps would be for the initiators of the more open and application independent data sharing initiatives to shift focus towards evaluating not only if data is available and shareable, but also actually used. While progress is being made in the right direction, as data ecosystems are set up in which data is indeed made available and shareable, there is a need to be mindful of actually getting past this intermediate maturity phase of having only a few pilot cases. Consequently, there are clear indications that the nuances mentioned in this paper must be taken into account by policymakers and public funding institutes when evaluating data ecosystems as a successful alternative to current data handling practices.

In regard to future research to be done, apart from what is mentioned in Section 5.1, the landscape of data ecosystems is expected to evolve. Therefore, this dataset should be updated on regular (for example yearly) basis. This must allow for a multi-year study in which it can be made clear whether there is an evolution in this landscape towards more open ecosystems actually reaching innovation potential as envisioned. A question to ask is if in time, more ecosystems with similar problems of a lack of data reuse pop up, or the landscape evolves to overcome this problem. One such an evolution is noticeable in a recurring report by the IDSA on their data ecosystems (International Data Spaces Association, 2024a). Every report, the amount of data ecosystems grow, but the operational ones remain limited. The idea is thus to continue this study in the future, and publish a yearly updated dataset. In case noteworthy evolutions are seen, a new report in the form of a research paper would be appropriate. Additional parameters could also be considered in future analyses, like active region, to allow for a systematic comparison between regions and policy regulations active in each region.

CRediT authorship contribution statement

Maarten de Mildt: Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Sofie Verbrugge:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization. **Didier Colle:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Acknowledgements

This research is supported by SolidLab Vlaanderen (Flemish Government, EWI and RRF project VV023/10); and imec ICON project SHARCS (Agentschap Innoveren en Ondernemen project nr. HBC.2022.0543).

Data availability

The original data accompanying this paper is available on GitHub at https://github.com/TechnoEconomics/DataEcosystemMarketAnalysis.

References

Askarzai, W., & Unhelkar, B. (2017). Research methodologies: An extensive overview. *International Journal of Science and Research Methodology*, *6*(4), 21.

Athumi. (2023). https://athumi.be/en/. (Accessed August 2023).

Baars, H., Tank, A., Weber, P., Kemper, H.-G., Lasi, H., & Pedell, B. (2021). Cooperative approaches to data sharing and analysis for industrial internet of things ecosystems. *Applied Sciences*, *11*(16), http://dx.doi.org/10.3390/app11167547, Retrieved from https://www.mdpi.com/2076-3417/11/16/7547.

Braud, A., Fromentoux, G., Radier, B., & Le Grand, O. (2021). The road to European digital sovereignty with Gaia-X and IDSA. *IEEE Network*, *35*(2), 4–5. http://dx.doi.org/10.1109/MNET.2021.9387709.

Bulger, M., Taylor, G., & Schroeder, R. (2014). Data-driven business models: Challenges and opportunities of big data.

Census. (2023). https://www.getcensus.com/. (Accessed August 2023).

Common European data spaces. (2024). https://digital-strategy.ec.europa.eu/en/policies/data-spaces. (Accessed June 2024).

Crepax, T., Gaur, M., & da Rosa Lazarotto, B. (2023). Measuring data access and re-use in the European Legal Framework for Data, from the General Data Protection Regulation (GDPR) law to the Proposed Data Act: the case of vehicle data. *Open Research Europe*, *3*.

Custers, B., & Bachlechner, D. (2017). Advancing the EU data economy: Conditions for realizing the full potential of data reuse. *Information Polity*, *22*(4), 291–309.

Custers, B., & Uršič, H. (2016). Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law*, *6*(1), 4–15. http://dx.doi.org/10.1093/idpl/ipv028.

da Rosa Lazarotto, B. (2022). The Data Act: empty promises for business-to-government data sharing? A critical analysis of the Proposal on the Data Act and its implications for the redistribution of data. *Paper Originally Published At Privacy in Germany (PinG) Datenschutz Und Compliance Journal*.

DataVaults Project (2024). DataVaults demonstrators. Retrieved from https://www.datavaults.eu/demonstrators/. (Accessed June 2024).

DataVaults.eu. (2023). https://www.datavaults.eu/. (Accessed August 2023).

de Mildt, M., Verbrugge, S., & Colle, D. (2023). Market analysis on data ecosystem initiators. http://dx.doi.org/10.5281/zenodo.8385400, Retrieved from https://github.com/TechnoEconomics/DataEcosystemMarketAnalysis.

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research Agenda. *Journal of Information Technology*, *33*(2), 113–124.

Digitaal Vlaanderen (2024a). Onze oplossingen - Open data. Retrieved from https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/open-data. (Accessed September 2024).

Digitaal Vlaanderen (2024b). Vlaamse digitale strategie. Retrieved from https://www.vlaanderen.be/digitaal-vlaanderen/vlaamse-digitale-strategie. (Accessed September 2024).

European Commission (2020). Towards a European strategy on business-to-government data sharing for the public interest. Retrieved from https://digital-strategy.ec.europa.eu/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more, Final report prepared by the high-level expert group on business-to-government data sharing.

European Commission (2024a). European blockchain services infrastructure (EBSI). Retrieved from https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/Home. (Accessed June 2024).

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

European Parliament and the Council (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Retrieved from http://data.europa.eu/eli/reg/2016/679/2016-05-04. (Accessed September 2024).

European Parliament and the Council (2022a). 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). Retrieved from http://data.europa.eu/eli/reg/2022/1925/oj. (Accessed September 2024).

European Parliament and the Council (2022b). 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). Retrieved from http://data.europa.eu/eli/reg/2022/868/oj. (Accessed June 2024).

European Parliament and the Council (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). Retrieved from https://eur-lex.europa.eu/eli/reg/2023/2854. (Accessed June 2024).

Farrell, E., et al. (2023). *European data spaces-scientific insights into data sharing and utilisation at scale: Tech. rep. no. JRC129900*, Joint Research Centre (Seville site).

Flemish smart data space. (2023). https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/vlaamse-smart-data-space. (Accessed August 2023).

Gaia-X. (2023). https://gaia-x.eu/. (Accessed August 2023).

Gaia-X lighthouse projects. (2023). https://gaia-x.eu/who-we-are/lighthouse-projects/. (Accessed September 2023).

Hardjono, T., & Pentland, A. (2019). Data cooperatives: Towards a foundation for decentralized personal data management. CoRR abs/1905.08819, Retrieved from http://arxiv.org/abs/1905.08819.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations and Production Management*, *36*(10), 1382–1406. http://dx.doi.org/10.1108/IJOPM-02-2014-0098.

Hilbig, R., Etsiwah, B., & Hecht, S. (2018). Berlin start-ups - the rise of data-driven business models.

HubSpot. (2023). https://www.hubspot.com/. (Accessed August 2023).

HubSpot (2024). Customer case study directory. https://www.hubspot.com/case-studies/directory. (Accessed June 2024).

iGrant.io. (2023). https://igrant.io/. (Accessed August 2023).

iGrant. io (2024). iGrant.io use case library. Retrieved from https://igrant.io/demo.html. (Accessed June 2024).

International Data Spaces Association (2024a). Data spaces radar. Retrieved from https://internationaldataspaces.org/adopt/data-spaces-radar/. (Accessed July 2024).

International data spaces association. (2024b). https://internationaldataspaces.org/. (Accessed June 2024).

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67–93, Retrieved from http://www.jstor.org/stable/249410.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660. http://dx.doi.org/10.1093/icc/dtaa002.

Kraemer, P., Niebel, C., & Reiberg, A. (2023). Gaia-X and business models.

Krämer, J., Shekhar, S., & Hofmann, J. (2022). Regulating algorithmic learning in digital platform ecosystems through data sharing and data siloing: Consequences for innovation and welfare. Retrieved from http://hdl.handle.net/10419/265645.
Lashkari, B., & Musilek, P. (2021). A comprehensive review of blockchain consensus mechanisms. *IEEE Access*, *9*, 43620–43652.
Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance. http://dx.doi.org/10.24251/HICSS.2021.733.
Meeco. (2023). https://www.meeco.me/. (Accessed August 2023).
Meeco (2024). Powered by Meeco. Retrieved from https://www.meeco.me/powered-by-meeco. (Accessed June 2024).
Mn, M., & Luna-Reyes, L. (2017). Open government data ecosystems: A closed-loop perspective. http://dx.doi.org/10.24251/HICSS.2017.327.
Mukhopadhyay, S., & Bouwman, H. (2019). Orchestration and governance in digital platform ecosystems: a literature review and trends. *Digital Policy, Regulation and Governance*, *21*(4), 329–351.
MyData Global (2024). MyData. Retrieved from https://mydata.org/. (Accessed July 2024).
OpenID Foundation (2024). OpenID Foundation. Retrieved from https://openid.net/. (Accessed June 2024).
OSLO standards. (2023). https://athumi.be/en/technologies/oslo-standards. (Accessed August 2023).
Osterwalder, A., Pigneur, Y., Papadakos, P., Bernarda, G., Papadakos, T., & Smith, A. (2014). *Value proposition design*. John Wiley & Sons.
Otto, B., ten Hompel, M., & Wrobel, S. (2022). *Designing data spaces*. Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-93975-5.
PubHubs (2024). PubHubs: Connecting ideas. Retrieved from https://app.pubhubs.net/. (Accessed July 2024).
Public spaces. (2023). https://publicspaces.net/english-section/. (Accessed August 2023).
Rufus Pollock (2011). Building the open data ecosystem. https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/.
Runeson, P., Olsson, T., & Linåker, J. (2021). Open Data Ecosystems — An empirical investigation into an emerging industry collaboration concept. *Journal of Systems and Software*, *182*, Article 111088. http://dx.doi.org/10.1016/j.jss.2021.111088, Retrieved from https://www.sciencedirect.com/science/article/pii/S0164121221001850.
S. Oliveira, M., Barros Lima, G. d. F., & Farias Lóscio, B. (2019). Investigations into data ecosystems: a systematic mapping study. *Knowledge and Information Systems*, *61*(2), 589–630. http://dx.doi.org/10.1007/s10115-018-1323-6.
Sambra, A. V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., et al. (2016). *Solid: a platform for decentralized social applications based on linked data*: *Tech. rep.*, MIT CSAIL & Qatar Computing Research Institute.
Sargent, T., & Denniston, L. (2023). Valuing data where are we, and where do we go next? CIGI Papers No. 280.
Schweihoff, J., Jussen, I., Stachon, M., & Möller, F. (2022). Design options for data-driven business models in data-ecosystems.
Schweitzer, H., & Metzger, A. (2023). Data access under the draft data act, competition law and the DMA: Opening the data treasures for cowmpetition and innovation? *GRUR International*, *72*(4), 337–356.
Shabani, M. (2021). The Data Governance Act and the EU's move towards facilitating data sharing. *Molecular Systems Biology*, *17*(3), Article e10229. http://dx.doi.org/10.15252/msb.202110229, Retrieved from https://www.embopress.org/doi/abs/10.15252/msb.202110229.
Snowflake. (2023). https://www.snowflake.com/en/. (Accessed August 2023).
Snowflake (2024). Healthcare and life sciences solutions. https://www.snowflake.com/en/solutions/industries/healthcare-and-life-sciences/. (Accessed June 2024).
Sorescu, A. (2017). Data-driven business model innovation. *Journal of Product Innovation Management*, *34*(5), 691–696. http://dx.doi.org/10.1111/jpim.12398.
Susha, I., Flipsen, M., Agahari, W., & de Reuver, M. (2020). Towards generic business models of intermediaries in data collaboratives: From gatekeeping to data control. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*: *vol. 12219*, (pp. 304–315).
Susha, I., van den Broek, T., van Veenstra, A. F., & Linåker, J. (2023). An ecosystem perspective on developing data collaboratives for addressing societal issues: The role of conveners. *Government Information Quarterly*, *40*(1), http://dx.doi.org/10.1016/j.giq.2022.101763.
The data governance act. (2023). https://digital-strategy.ec.europa.eu/en/policies/data-governance-act. (Accessed September 2023).
The digital markets act. (2023). https://digital-markets-act.ec.europa.eu/index_en. Accessed September 2023.
Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives*. OECD.
van Dijck, J., & Jacobs, B. (2024). PubHubs: A secure and trustworthy online platform. Retrieved from https://www.pubhubs.net/. (Accessed June 2024).
Van Lancker, D., Colpaert, P., Delva, H., Van de Vyvere, B., Meléndez, J. R., Dedecker, R., et al. (2021). Publishing base registries as linked data event streams. In M. Brambilla, R. Chbeir, F. Frasincar, I. Manolescu (Eds.), *Web engineering* (pp. 28–36). Cham: Springer International Publishing.
Van Ooijen, I., & Vrabec, H. U. (2019). Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. *Journal of Consumer Policy*, *42*, 91–107.
Verstraete, M., Fierens, M., Verbrugge, S., & Colle, D. (2023). Assessing the economic viability of data intermediation services as defined in the DGA. In *Proceedings 62nd international FITCE congress, "services and infrastructures toward smart society"*.
Vlaamse overheid (2024). Vlaamse smart data space portaal: Use cases. https://www.vlaanderen.be/vlaamse-smart-data-space-portaal/use-cases. (Accessed June 2024).
von Ditfurth, L., & Lienemann, G. (2022). The Data Governance Act: – Promoting or restricting data intermediaries? *Competition and Regulation in Network Industries*, *23*(4), 270–295. http://dx.doi.org/10.1177/17835917221141324.
Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems*, *4*(2), 74–81. http://dx.doi.org/10.1057/ejis.1995.9.
Wernick, A., Olk, C., & von Grafenstein, M. (2020). Defining data intermediaries. http://dx.doi.org/10.26116/techreg.2020.007.
Yin, R. K. (2018). *Case study research and applications*. CA: Sage Thousand Oaks.
Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*(1–2), 17–33.