

de Mildt, Maarten; Verbrugge, Sofie; Colle, Didier

## Conference Paper

# The reality of data-driven ecosystems: how do theoretic benefits hold true in practice?

32nd European Conference of the International Telecommunications Society (ITS):  
"Realising the digital decade in the European Union – Easier said than done?", Madrid,  
Spain, 19th - 20th June 2023

### Provided in Cooperation with:

International Telecommunications Society (ITS)

*Suggested Citation:* de Mildt, Maarten; Verbrugge, Sofie; Colle, Didier (2023) : The reality of data-driven ecosystems: how do theoretic benefits hold true in practice?, 32nd European Conference of the International Telecommunications Society (ITS): "Realising the digital decade in the European Union – Easier said than done?", Madrid, Spain, 19th - 20th June 2023, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/277952>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The reality of data-driven ecosystems: how do theoretic benefits hold true in practice?

Maarten de Mildt<sup>1\*</sup>, Sofie Verbrugge<sup>1</sup>, Didier Colle<sup>1</sup>

<sup>1</sup>IDLab, Ghent University, Ghent, Belgium

\*Corresponding author: Maarten de Mildt; [maarten.demildt@ugent.be](mailto:maarten.demildt@ugent.be)

## Abstract

The concept of data-driven ecosystems is proposed as a solution to manage increasingly complex data processes. A common narrative 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, it is not standard practice yet. A critical look at the way the market is implementing this concept was needed. A market analysis was performed, linking ecosystem types to benefits expected in theory, and offered in practice by ecosystem initiator services. Three main ecosystem types, based on their scope, were identified and analysed: 1) single-organisation ecosystems, 2) application domain specific ecosystems and 3) cross-domain ecosystems. This revealed two main insights. On the one hand, achieving the innovation potential remains a challenge, especially for more broadly scoped ecosystems. On the other hand, the services offered in the more broadly scoped ecosystems, do support this potential. Therefore, providing an answer to the question "how do theoretic benefits hold true in practice?".

**Keywords:** data-driven ecosystem; data ecosystem; ecosystem initiator; benefit analysis; value offering; market analysis

## 1 Introduction

The demand for data and accompanying services is ever-increasing, which poses several challenges in managing data driven systems. In response, the concept of data ecosystems has been examined as a facilitator of managing data-driven processes (Marcelo, Barros Lima, & Farias Lóscio, 2019; de Reuver, Sørensen, & Basole, 2018; Susha, van den Broek, van Veenstra, & Linåker, 2023). A common narrative is that an increased amount of data will be available, of higher quality, and this will lead to an increase in innovation. This paper aims to deliver an overview of the adoption of the data ecosystem concept, and the challenges its adoption is facing.

This evolution of data ecosystems, is one of the responses to a larger (r)evolution on the Web. More data is being generated, more data is being used, processes are becoming more complex, and legislation is adapting to these changes (e.g. the Data Governance Act regulating digital platforms (von Ditfurth & Lienemann, 2022)). In general, 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 do control large amounts of data become increasingly powerful. These are believed to be clear incentives to try and change the way data is handled.

Data ecosystems propose a solution to these problems in which data, contributions to it and management of it, is distributed over all involved parties. Essentially, this is expected to lead to, among others, 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 benefits considered in this paper, which will be explained in section 5:

- More available data
- Improved data quality
- Increased innovation

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

Table 1: Several definitions of data ecosystems in related work, including our working definition. These were gathered from technical, socio-technic and economic literature. Although differing views, this table shows similarities are to be found.

## 2 Problem statement

Even though these benefits seem promising at first, barriers like the complexity of co-operation tasks (e.g. deciding on common standards) also prevent market-adoption of this ecosystem concept. Therefore, these benefits do not always seem to reach their potential in reality. A critical look at the way the market is implementing this concept, and how this influences benefits offered, is needed. In order to tackle this problem, several questions must be answered:

- Which benefits do actually ecosystems offer, and to whom?
- Who provide the services that offer these benefits?
- What are the challenges these ecosystem service providers face?
- Why are these ecosystems not reaching their potential yet?

## 3 What are data ecosystems?

**Related work on Data ecosystems** The first research question introduced in the previous section, was "what are data ecosystems?". Given the concept is not a novel one, this has lead to different definitions popping up, from different perspectives. An overview of the most common ones is listed in table 1. 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 et al., 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 organizational processes and standards" (sociotechnical view). Again, emphasizing that multiple perspectives on these concepts exist.

**Definition** Using the common elements in several of these definitions, data-driven 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.

## 4 Methodology

As the title of this paper suggests, the proposed methodology has both a theoretical and a practical side, shown in fig. 1. In order to evaluate if theory and practice match, regarding ecosystem benefits, there is a need for an evaluation method that combines both. The perspective taken here, is a "theory-first" one, meaning that literature will provide the necessary tools in order to evaluate the practical implementations. The following paragraphs describe both individual analyses, and their interaction, which must make them comparable.

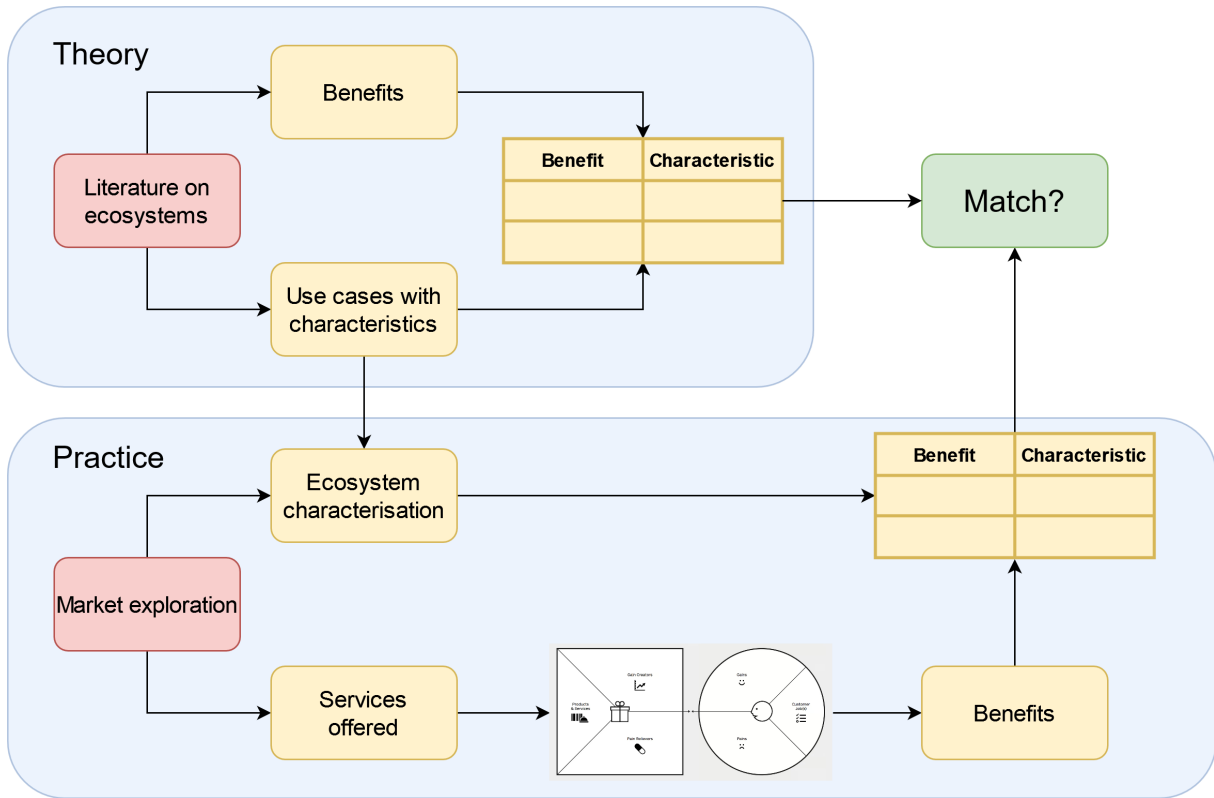


Figure 1: Methodology used showing both a theoretic and practical analysis, and how they can be compared on equal terms. The theoretic analysis stems from a literature deep-dive. This literature typically researched a few use cases to induce general benefits. The characteristics (e.g. open) of these use cases can be re-used in the practical analysis to determine what type of ecosystems are set up by the market. After doing a Value Proposition Design analysis of the services offered in the market, we achieve a similar overview in both analysis of benefits linked to characteristics.

**Theoretical analysis** As mentioned, a multitude of studies have been performed on data ecosystems. Each assign several benefits to the approach, depending on the angle taken. Simultaneously, they typically use case studies to either provide input or validation. Therefore, a link can be made between both, in a straightforward manner. For example, if a benefit found is that more data is contributed in ecosystems, and the case studies used were open data ecosystems, then a benefit that can be linked to open ecosystems, is increased data availability.

**Practical analysis** When it comes to performing a practical analysis, however, there is a need for a framework to evaluate this consistently. Two main sources of information can typically be found when exploring the data ecosystem market. On the one hand, information on what service they offer, which leads to their value propositions. On the other hand, background information on how these services are implemented, and what role the initiator plays therein, as well as for whom they implement ecosystem services. Thus providing the information needed to characterise the ecosystems they set up.

When it comes to extracting benefits, an additional step is needed. In order to go from value offering, to actual benefits, the methodology used is Value Proposition Design (Osterwalder et al., 2014). Value Proposition Design zooms in on the value offering for specific customer segments of a business model. Therefore, it is possible to find out whether the values offered by a certain product or service actually solve the problems or create the opportunities the customer is looking for in his tasks. Given this methodological framework that allows a practical analysis of services offered in the market, and extracting benefits (solving problems and creating opportunities), this is deemed a suitable candidate to analyse the practical side of the methodology.

**Matching theory and practice** Finally, this delivers both a theoretic overview of benefits and characteristics, and a practical one. Through the similar characteristics, they can then be compared easily. This highlights where there is a match between ecosystem design decisions and envisioned benefits, and where there is a mismatch.

## 5 Theoretic analysis

In the theoretic analysis, both a benefit study and a link with the use cases used to deduct these benefits, is made. These use cases carry certain characteristics, which then can be linked to these benefits. Next to this link, the benefits explained in the introduction are clarified. Both steps of the theory side in the presented methodology (fig. 1) are thereby covered in this section.

### Ecosystem Characteristics: Openness, Data access and Initiator Goal

Three main ecosystem characteristics are considered: Openness, Data access and Initiator Goal, defined in the following sections. The perspective taken for these characteristics was that of the initiator of an ecosystem, who is the decision maker here. Next to this, they had to be applicable to ecosystems, and not specific use cases within the ecosystem. This means that no matter the use case analysed in a specific ecosystem, these characteristics should not change. Fig. 2 shows an overview in the table listed as "Ecosystem Characteristics".

**Openness** Openness of the ecosystem represents the possibility for new actors to participate in the ecosystem that are not part of predefined processes, or known when the ecosystem was set up. Initiators thus have to make the decision whether they open up their ecosystem. (Mukhopadhyay & Bouwman, 2019) Describe membership openness as both being open or closed to multiple (external) individuals and firms. Open ecosystems are often represented by open data initiatives (Lis & Otto, 2021), by letting any actor contribute data freely. (Zuiderwijk, Janssen, & Davis, 2014) Describe the open data ecosystem as loosely interconnected participants in self-organizing 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. (Marcelo 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 & Bouwman, 2019). The alternative being a closed ecosystem. Only actors present at the time of establishing the ecosystem are then allowed to participate.

**Data access** Data access defines dependency of actors on certain applications to access, share and re-use data. An initiator has to decide if all actors can access data and/or exercise certain rights on the data that flows in the ecosystem. (Lis & 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 & Bouwman, 2019) in regard to platform leaders (here: initiators) and its complementers (here: all other ecosystem actors). If data is siloed (Krämer, Shekhar, & Hofmann, 2022), it is deliberately kept within an application's functioning, meaning there is typically no way to extract or share data to be used in other applications. On the other hand, if actors are allowed to excerpt authority over ecosystem data independent of applications, an application independent ecosystem policy is created. This fits the description of Personal Data Ecosystems in (Marcelo et al., 2019), in which actors control their data. At this point, actors should be able to extract data independent of the tools they used to generate the data, giving them the freedom to choose applications of their liking.

**Initiator goal** 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. In the first case, 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. The latter initiator goal of privacy protection is typically centred around public values, rather than economic gains.

### Ecosystem benefits and link to characteristics

As mentioned, many benefits are linked to an ecosystem approach, of which the common narrative is that an increased amount of data will be available, of higher quality, and this will lead to an increase in innovation (Marcelo et al., 2019; Susha et al., 2023; Krämer et al., 2022). While this sounds promising, the ecosystem approach is far from being the golden standard. The next paragraphs will clarify these benefits, and link them to the characteristics of the previous section. Fig. 2 shows an overview in the table listed as "Ecosystem benefits and characteristics". Each of these characteristics have more (specific) benefits, as well as links to other benefits, but for brevity, the key takeaways are listed here.

**Open ecosystems improve data quality** An open ecosystem leads to an improvement in data quality (Susha et al., 2023; Marcelo 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. Therefore, incorrect data can be called out, and the

responsible parties are present in the ecosystem to act on it. 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. Essentially leading to more verified and up-to-date data, which is considered as 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; Marcelo 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 order to be able to share data, actors need to speak a common language in the form of standards. While wanting to share data leads to more standardisation, more standardisation also leads to more data being shareable. Finally, if applications do not limit their users, (more) people are empowered to contribute (more) data ("data democracy").

**Data sharing brings forth innovation** Thirdly, data sharing must lead to innovation (Marcelo 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 fund allocation shifts explain this benefit. When breaking down silo's, sharing data, there is less competitive advantage of having done extensive data mining. This might cause a shift towards the more innovative services to win over the market. 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. Finally, if, for example, data mining is not a key activity for organisations, they can re-allocate these funds.

## 6 Practical analysis

The practical analysis, referring to the methodology of fig. 1 as "practice", consists of several steps that structure this analysis. First, a general market exploration lead to three main types based on the ecosystem scope. For each scope, several examples are characterized and their value offering is analysed using Value Proposition Design. From these examples, general benefits and conclusions can then be drawn to compare them to the theoretical analysis.

### Data ecosystem types based on scope

Given the broadness of the ecosystem concept, there are diverse ways of implementing data ecosystems. The ecosystem archetypes, identified in (Susha et al., 2023), confirm this view. Some approaches considered by them are: 1) a platform-centric approach, 2) a keystone-centric approach and 3) an intermediary-based system. These determine the organizational structure of the ecosystem, and who the initiating actors (called conveners by (Susha et al., 2023)) are. One of their insights, crucial for this paper, is that the implementation influences the roles of, and benefits offered by, these initiators of ecosystems, forming the basis of the problem statement of the practical side of this research.

**Ecosystem scopes: Single organisation, Application domain Specific, Cross-domain** Three main types of ecosystems are considered here, based on the scope of the ecosystem. This scope defines the boundaries of the ecosystem regarding who are participating actors and who are considered outsiders. The three scopes considered, shown in fig. 3, are: 1) Single organisation scope, 2) Application domain specific scope, 3) Cross-domain scope. First, there are the ecosystems connecting a single organisation. Such ecosystems are typically confined within an organisation's walls, sharing data across departments. Secondly, the application domain specific ecosystems target industry verticals within application domains. They promote co-operation outside of organisation boundaries, albeit within specific domains. Finally, the cross-domain ecosystems represent a more divergent set of ecosystems. Such ecosystems are typically centred around data intermediaries, relating to the intermediary-based ecosystems of (Susha et al., 2023), who facilitate the sharing of data. Ideally, these alleviate the potential of each piece of data and re-use it for as many use cases as possible.

### Market exploration

A market exploration was done, representing all three scopes, in order to get an overview of the existing ecosystem services. Some examples are picked here whom are representative for the market are discussed in more detail. Given the broadness of this market exploration, different sources were used. These reflect our interpretations based on meetings held with representatives, events attended, co-operations in research projects, and documentation like whitepapers.

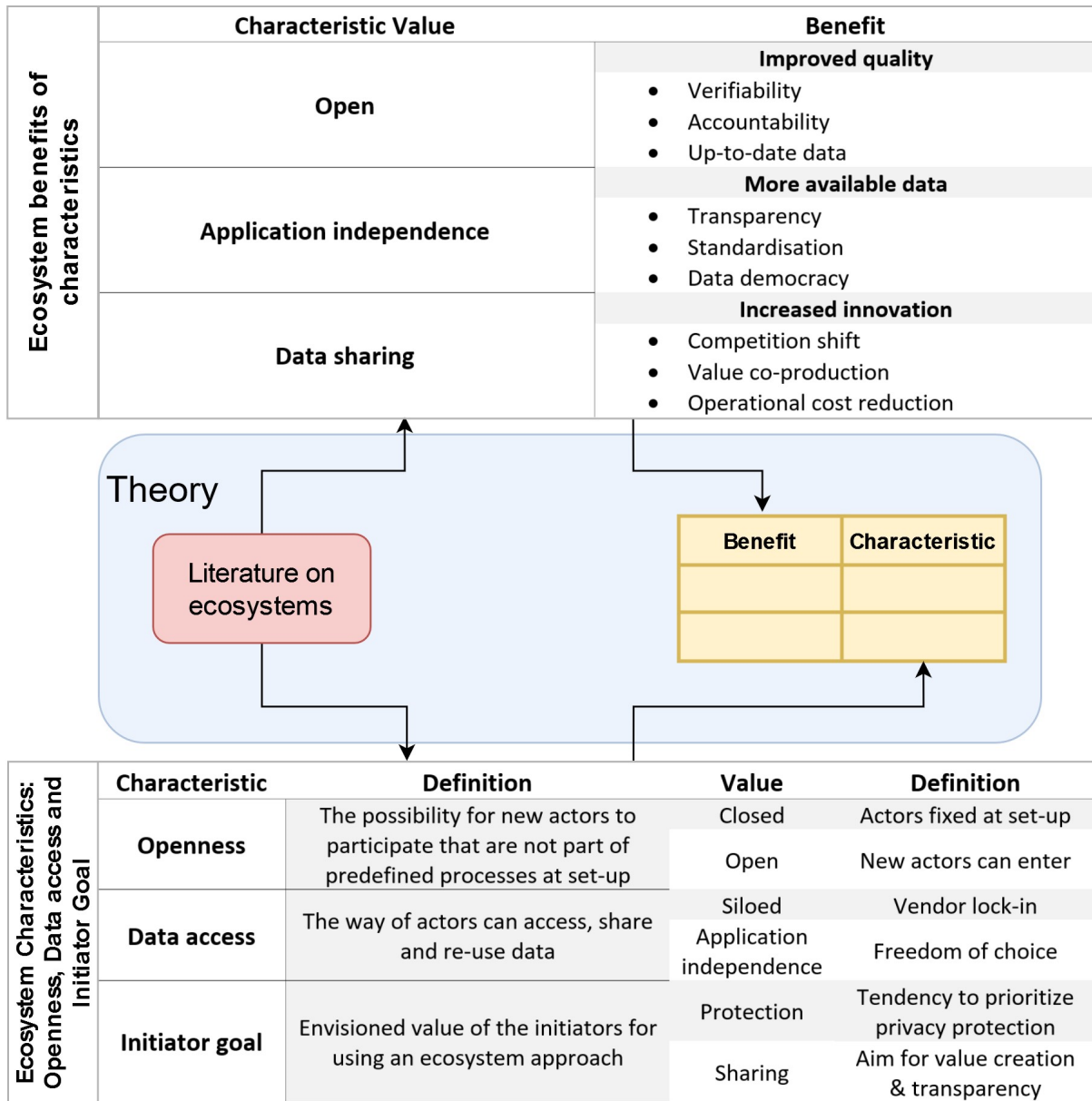


Figure 2: The theoretical analysis matches benefits typically attributed to data ecosystems to values taken by data ecosystem characteristics. The top part represents the "benefits" step of the methodology. This table represents a common narrative found in literature that open (regarding participation) ecosystems lead to higher quality data, application independence (tearing down silos) leads to more available data, and a goal of sharing data leads to more innovation. Note that specific reasons are listed below each general benefit. The bottom part shows considered ecosystem characteristics and their accompanying values. These are deduced from several examples found in literature, representing step "use cases with characteristics" of the methodology.

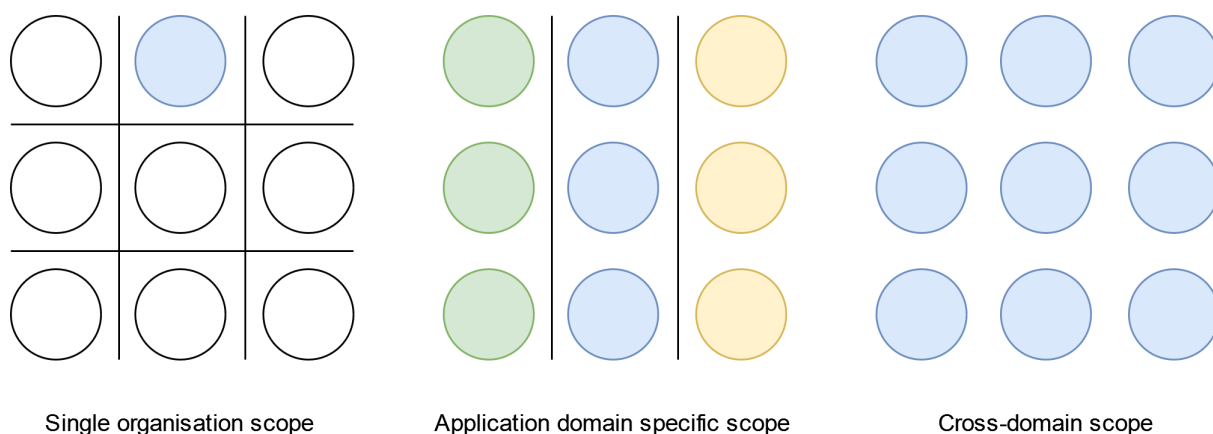


Figure 3: Different ecosystem types initiated by the market, based on their target audience, leading to differing scopes. First, the single organisation scope, typically connecting silos within an organisation’s walls. Secondly, the application domain specific ecosystems, who target industry verticals by sharing data with domain partners. Finally, the cross-domain ecosystems represent a more divergent set of ecosystems. Such ecosystems are typically centred around maximizing the re-use potential of data, through data intermediaries.

**Single organisation scope** Within the single organisation scope the examples that will be discussed are Snowflake<sup>1</sup>, HubSpot<sup>2</sup> and Census<sup>3</sup>. All three of these offer services that focus on connecting different departments through data, within an organisation. Essentially, they integrate an organisation’s data silos into the different applications they use. An overview of the analysis on these examples is given in table 3.

**Application domain specific scope** Analysing the application domain specific scope, some initiatives that have gained attention in recent years are Gaia-X<sup>4</sup>, the industry vertical cases of Athumi<sup>5</sup> and Public Spaces<sup>6</sup> supported initiatives like PubHubs<sup>7</sup>. These parties focus on orchestration of ecosystems, stimulating innovative initiatives and connecting parties, while providing support or regulatory backing. This could be through setting policies, auditing if these policies are respected, promoting standards and even providing infrastructure. One of the key aspects of such an ecosystem is that each step in such an industry vertical is represented by multiple actors, promoting interoperability between each of them. An overview of the analysis on these examples is given in table 4.

**Cross-domain scope** Finally, looking at the cross-domain scope, a common denominator for all the ecosystems in this space is there being no predefined use case or sector in which the data is to be used. The boundaries of the ecosystem are typically not based on a certain industry or an organization, but rather the premise of maximizing shared data. Both by sharing the same data with multiple parties, but also building upon that data to create new data, for new purposes. Some of the examples included centre around the individual like DataVaults.eu<sup>8</sup>, Meeco<sup>9</sup> and iGrant.io<sup>10</sup>. On the other hand, the Flemish Smart Data Space<sup>11</sup> is also included, as an ecosystem that is built around intermediaries that facilitate data sharing, targeting B2B interactions mainly. An overview of the analysis on these examples is given in table 5.

## Ecosystem initiator services analysis

A more in-depth analysis of the example companies that represent each ecosystem scope is discussed in the next sections. For each company, their characterization based on the theoretic analysis, and an overview of the services they offer is given. In order to provide structure in each company’s own interpretation of certain services, these will be linked to common service types found in literature on data-driven business models, shown in table 2. When listing services in the following sections, this is a reference to table 2. Finally, the results of conducting a Value Proposition Design analysis of each example, revealing benefits these services result in, are assessed.

<sup>1</sup><https://www.snowflake.com/en/>, Accessed April 2023

<sup>2</sup><https://www.hubspot.com/>, Accessed April 2023

<sup>3</sup><https://www.getcensus.com/>, Accessed April 2023

<sup>4</sup><https://gaia-x.eu/>, Accessed April 2023

<sup>5</sup><https://athumi.be/en/>, Accessed June 2023

<sup>6</sup><https://publicspaces.net/english-section/>, Accessed April 2023

<sup>7</sup><https://pubhubs.net/en/>, Accessed April 2023

<sup>8</sup><https://www.datavaults.eu/>, Accessed May 2023

<sup>9</sup><https://www.meeco.me/>, Accessed May 2023

<sup>10</sup><https://igrant.io/>, Accessed May 2023

<sup>11</sup><https://athumi.be/en/technologies/flemish-smart-data-space>, Accessed June 2023



<b>Service</b>	<b>Description</b>	<b>Sources</b>
<b>Consultancy</b>	General advisory services, provisioning of expertise knowledge on data-driven processes.	(Bulger et al., 2014; Susha et al., 2023)
<b>Infrastructure</b>	Provisioning of hardware components and their management, as well as dedicated technology integrations onto such hardware components.	(Bulger et al., 2014; Susha et al., 2023; Kraemer, Niebel, & Reiberg, 2023; Braud, Fromentoux, Radier, & Le Grand, 2021)
<b>Matchmaking</b>	Connecting actors, matching demand and supply.	(Susha et al., 2023; Kraemer et al., 2023)
<b>Auditing</b>	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)
<b>Software-as-a-Service</b>	Software delivery on-demand, provisioning of (pluggable) components meant to perform specific data processing aspects or functionalities.	(Hilbig, Etsiwah, & Hecht, 2018; Sorescu, 2017)
<b>Regulation compliance</b>	Unburdening or automation of regulatory compliance.	(Kraemer et al., 2023; Braud et al., 2021; <i>Designing Data Spaces</i> , 2022)
<b>Data standardisation</b>	Enhancing data using metadata, modelling data, and adapting data to standards to ease communication.	(Susha et al., 2023; <i>Designing Data Spaces</i> , 2022)
<b>Access granting mechanisms</b>	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, & de Reuver, 2020; Schweihoff, Jussen, Stachon, & Möller, 2022)
<b>Data-as-a-service</b>	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)
<b>Data analytics</b>	Services that generate and derive insights, advice, and visualizations.	(Bulger et al., 2014; Sorescu, 2017; Susha et al., 2020; Hartmann et al., 2016)
<b>Clearinghouse</b>	A middleman service that approves data exchanges according to policies, e.g., are access rights to data respected.	(Susha et al., 2020; Wernick, Olk, & von Grafenstein, 2020)
<b>Data broker</b>	Monetizing 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)
<b>Governance tools</b>	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; <i>Designing Data Spaces</i> , 2022; Susha et al., 2020)

Table 2: Overview of common data services that apply to data ecosystems. Specifically, these are typically offered by the initiator(s) of the ecosystem, the actor(s) in the ecosystem that create and typically facilitate the development of the ecosystem. The, possibly indirect, beneficiaries should be all ecosystem actors.

Example	Openness	Data Access	Initiator Goal	Services
<b>Snowflake</b>	Closed	Siloed	Data sharing	Consultancy - Infrastructure - SaaS - Data standardisation - Data-as-a-service - Data analytics - Governance tools
<b>HubSpot</b>	Closed	Siloed	Data sharing	SaaS - Data-as-a-service - Data analytics
<b>Census</b>	Closed	Siloed	Data sharing	SaaS - Data-as-a-service - Data analytics

Table 3: Characterized examples and their services offered in the single organisation scope

### Single organisation scope

**Snowflake** Snowflake 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, apart from their marketplace which is a different service, connecting data departments with departments like marketing. Therefore, their ecosystems are closed, the data is still stored in silos and only a front-end view like an API connects these silos. However, Snowflake looks for clients that want to share data, albeit within limited organizational or partnership boundaries.

**HubSpot** HubSpot 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. Again aiming their products towards more closed ecosystems, within company boundaries, and connecting silos (often simply syncing data stores rather than using a new central point). Again, sharing data within their organisation.

**Census** A similar offering is given by Census, 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. This means they typically set up closed ecosystems. Next to this, they build mappers that do not fundamentally change the paradigm on how data is stored in silos, but connects them. Finally, the goal of census is to make their customers able to share data they previously had a hard time transferring between departments.

**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. Therefore, data becomes re-usable and co-production of value using data and knowledge from different teams is made possible. Optimization of processes, by making sure there is a common data source that is re-usable, 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. Summarized, 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.

### Application domain scope

**Gaia-X** Gaia-X is an international, European, 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. One of their main services offered is their 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.

**Athumi** One technology that allows data ecosystems to be developed is that of Solid ([Sambra et al., 2016](#)). Athumi 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. Providing citizens with a way to share their diploma, for example, they create an open

Example	Openness	Data Access	Initiator Goal	Services
<b>Gaia-X</b>	Open	Both	Data sharing	Matchmaking - Auditing - Data-as-a-service
<b>Athumi</b>	Open	Application independence	Data sharing	Infrastructure - Auditing - Data standardisation - Access control - Data-as-a-service
<b>Public Spaces</b>	Open	Application independence	Privacy protection	Consultancy - Auditing

Table 4: Characterized examples and their services offered in the application domain specific scope

ecosystem. Given they make use of Solid, the essential idea is to use personal data stores for each citizen. This means they fundamentally break open (government) silos and redistribute this data across the ecosystems they build. Therefore, Athumi tries to jumpstart new economies by facilitating data sharing.

**Public Spaces** In the Netherlands, an initiative was created that focusses on public values, and is therefore a bit atypical from the earlier mentioned organisations. One example supported by Public Values is that of PubHubs. 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. As public values are central here, the ecosystems they facilitate are open and based on breaking down silos and redistributing data across the ecosystem.

**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. 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. Summarized, regarding the benefits named in the theoretic analysis: additional 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. It must be noted that no conclusion is made on if this is a negative or positive strategy, but rather a neutral observation.

### Cross-domain scope

**DataVaults** The DataVaults.eu 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 monetization streams.

**Meeco** Meeco 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.

**iGrant** iGrant.io 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.

**Flemish Smart Data Space** Finally, the Flemish Smart Data Space 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

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

Table 5: Characterized examples and their services offered in the cross-domain scope

Streams (Van Lancker et al., 2021). An important note is that Athumi offers their services in the Flemish Smart Data Space, showing an initiator can support multiple kinds of ecosystems.

**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 anonymizing 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. Summarized, 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. This is a neutral stance again, but given the complexity of this data sharing, without clear limitations, this is a challenging one. 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.

## 7 Do theoretic benefits hold true in practice?

Having analysed both theory and practice, we now have an overview of the characteristics and their link to benefits for both. First, the theoretic benefits that can be linked to the observed characteristics for the different examples, per scope, are listed. Secondly, the benefits that were extracted from the services offered by the market are listed in similar fashion. Finally, the final subsection is devoted to a discussion on whether they match, and to add some nuance to the mismatches. Fig. 4 depicts all three aspects.

### Theoretic benefits linked to characteristics perceived in practice

**Single organisation scope** Regarding the single organisation scope, the following characteristics were consistently observed: 1) Openness: closed, 2) Data access: siloed and 3) Initiator goal: data sharing. Essentially, the only benefit considered here, these should offer is thus increased innovation due to data sharing if we link back to the theoretic benefits from section 5.

**Application domain scope** The application domain held the following characteristics: 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. However, the nuance must be added here that the Public Spaces does not aim to innovate, and Gaia-X does not meddle with data storage, not forcing data access independent of applications. Nonetheless, there is still a dominance of characteristics that should deliver all three benefits.

**Cross-domain scope** Finally, when looking at the cross-domain scope there is a clear consensus on innovators building these ecosystems: 1) Openness: open, 2) Data access: application independence and 3) Initiator goal: data sharing. Therefore, all three benefits should be perceived in practice. While this seems similar to the

application domain specific scope in benefits, it is notable that there seems to be a common view on how the ecosystem is set up for the cross-domain scope.

## Practical benefits extracted from services perceived in practice

**Single organisation scope** For the single organisation scope the observed services consistently contained data analytics. Specifically, these focussed on making use of this now combinable data to bring forth novel insights, which must lead to the benefit of increased innovation. Secondly, due to there now being a single-source-of-truth among departments, the quality of the data also improves.

**Application domain scope** The application domain specific ecosystems do 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 service providers. An important nuance to be made here, is that they do offer consultancy services, which is not directly offering this benefit, but supporting it.

**Cross-domain scope** In the cross-domain ecosystems, the focus of the offered services is on supporting or intermediary 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 service providers. However, contrary to the application domain specific scope, there is no clear supporting party. Orchestration of the ecosystem is more of an organic process in the form of a distributed network of data intermediaries.

## Match theory and practice

The final step is now to put the theoretic and practical analysis tables next to one another and compare. The central table in fig. 4 shows the result of this comparison, with benefits observed (checkmarks) and not observed (cross-marks) in theory and practice (respectively) pictured. Two similar symbols therefore denote a match. In the next paragraphs, some key takeaways that this matching displays are discussed.

- **The single organisation scope is the only type where initiators offer innovative services:** While theory expects all ecosystem approaches to innovate in the end, only the single organisation scope offers services that gather new insights from data by 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, focussing on initiating and supporting the ecosystem. This reveals a mismatch between theory, expecting innovation, and practice, where most initiators do not implement the innovative services themselves.
- **Application domain specific and cross-domain ecosystems carry innovation potential:** Both ecosystem types offer services that carry the benefits of making more data, of higher quality, available. Therefore, (additional) data both becomes accessible and re-usable for the actors in the ecosystem, compared to before adopting an ecosystem approach. This indicates the currently yet often unachieved (regarding the previous takeaway) innovation potential is represented by the market, in line with theoretic expectations.
- **Application domain specific ecosystems are working towards closing this innovation gap:** The application domain specific initiators typically offer or support (the development of) matchmaking services between data providers and service providers. This is represented by governance tool services, for example the federation toolbox by Gaia-X. Innovative services are kickstarted by these supporting services, showing this innovation gap between theory and practice might be closed in the future. While the evolution between ecosystem scopes was not examined in this paper, this might indicate this ecosystem type represents a pragmatic approach in an evolution from single organisation scopes to cross-domain scopes.
- **Cross-domain ecosystems face the biggest challenges on making data shareable:** While cross-domain ecosystems might carry the most potential, endless data sharing across use cases and domains introduces additional actors to the ecosystem. Each of these must adapt to the agreed-upon standards in the ecosystem, might require additional trust mechanisms to provide data to non-industry partners who they know less well, might not know beforehand what data is available or which purpose this data can serve for them... This is reflected in the shift of value offering by the initiators of these ecosystems, which focus on services that aim to provide, among others, trust and interoperability.
- **The single organisation scope offers improved quality, although it is a closed ecosystem:** Surprisingly, this analysis also showed a mismatch in a positive sense. The single organisation ecosystem service providers actually offer services that improve the quality of data, even though it is a closed ecosystem. This contrasts with literature in a way that the reasons given for this improved quality (being verifiability, accountability and up-to-date data) are not typically believed to be present in closed ecosystems. However,

the reality is that most organisations are still experiencing difficulties to communicate between different departments and subsidiary companies. Consequently, these services do offer them these benefits.

## 8 Conclusions and future work

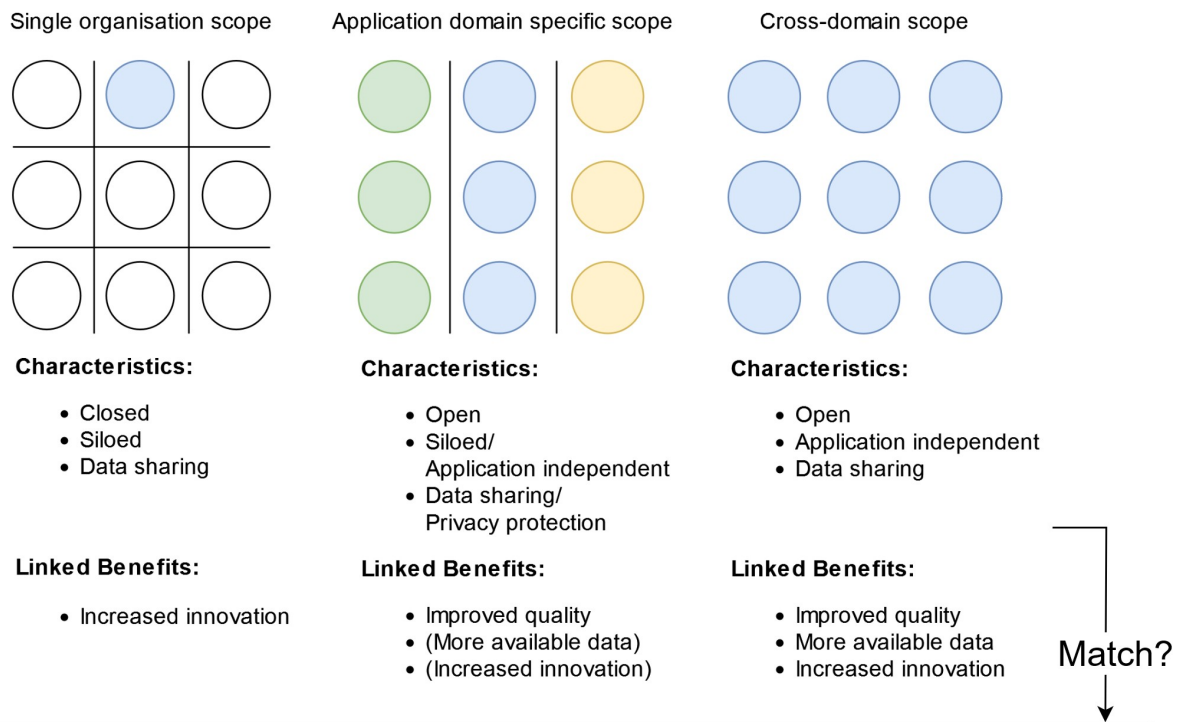
In this research, an exploration of different types of data ecosystems, and initiator's services therein, was performed. This was linked to a study of the narrative found in literature that open, application independent ecosystems aimed at data sharing must deliver more data of higher quality, and lead to innovative services. In order to do extract benefits in this market exploration, Value Proposition Design was used to analyse the services offered by several market representatives. From this combination of a theoretic and practical analysis, a match and mismatch was discussed by putting both next to one another for the different types of ecosystems observed.

**Achieving the innovation potential remains a challenge** A first main conclusion is that not all benefits are put into practice yet, specifically the one of increased innovation. While steps are being taken in the right direction, the potential for innovation is yet to be realized 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. It is often left to the service providers to deliver innovative services, although advisory services are found in the market to guide them in this process. This is a neutral take on the ecosystem approach, as the initiators might try to let the ecosystem evolve on its own. It must be emphasized that there is no position taken here on whether this approach works or not, it is simply the observation made.

**The broader the scope, the bigger the challenge** Another important takeaway is that, the more broadly scoped the ecosystem, the greater the discrepancy between theoretic potential benefits and benefits employed in practice. While the single organisation scoped ecosystems offer services that actually turn data into insights, the cross-domain ecosystems are still in a preliminary phase where the focus is on making data shareable, rather than actionable. Finally, the application domain specific ecosystems often remain neutral here, but provide some stimulating tools like a platform to connect data- and service providers.

**Future work** There are many challenges still to be analysed in more detail. Continuing this market analysis, and expanding on the characteristics used to analyse them, is to be tackled in future work. This must allow for the methodology proposed here to be validated as a general framework for analysing ecosystems. Both before set-up, and as an evaluation for operational ones. Next to this, an extensive study on the business models for ecosystem initiators is to be made, as these were only analysed on a value offering level for now.

### Theoretic benefits linked to characteristics perceived in practice



Scope \ Benefit	Improved quality	More available data	Increased Innovation
Single organisation	X✓	XX	✓✓
Application domain specific	✓✓	✓✓	✓X
Cross-domain	✓✓	✓✓	✓X

### Practical benefits extracted from services perceived in practice

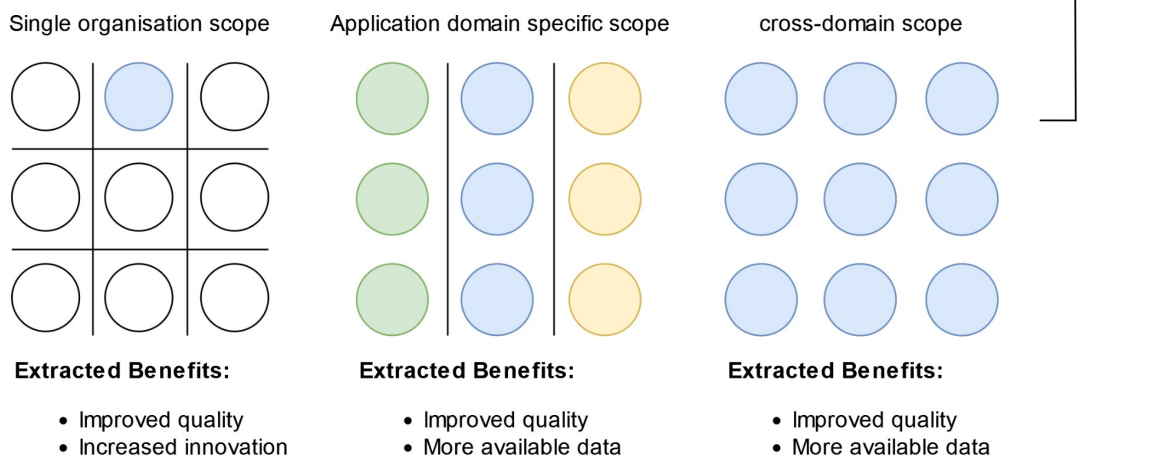


Figure 4: Benefits observed (checkmarks) and not observed (cross-marks) in theory and practice (respectively). Two similar symbols therefore denote a match. The single organisation scope is the only type in which the initiators offer services that increase innovation in practice, contrary to what theory preaches. However, application domain specific and cross-domain ecosystems carry innovation potential, and through targeted supporting services the application domain specific ecosystem initiators are making efforts towards closing this gap between theory and practice.

## Acknowledgements

This research is supported by SolidLab Vlaanderen (Flemish Government, EWI and RRF project VV023/10).

## References

- 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). Retrieved from <https://www.mdpi.com/2076-3417/11/16/7547> DOI: 10.3390/app11167547
- 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. DOI: 10.1109/MNET.2021.9387709
- Bulger, M., Taylor, G., & Schroeder, R. (2014). Data-driven business models: Challenges and opportunities of big data..
- de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, *33*(2), 124–13.
- Designing data spaces.* (2022). Springer International Publishing. DOI: 10.1007/978-3-030-93975-5
- 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. DOI: 10.1108/IJOPM-02-2014-0098
- Hilbig, R., Etsiwah, B., & Hecht, S. (2018). Berlin start-ups - the rise of data-driven business models.
- 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>
- Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance.. Retrieved from <https://doi.org/10.24251/HICSS.2021.733> DOI: 10.24251/HICSS.2021.733
- Marcelo, M. I., 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. Retrieved from <https://doi.org/10.1007/s10115-018-1323-6> DOI: 10.1007/s10115-018-1323-6
- Mn, M., & Luna-Reyes, L. (2017, 01). Open government data ecosystems: A closed-loop perspective.. DOI: 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.
- Osterwalder, A., Pigneur, Y., Papadakos, P., Bernarda, G., Papadakos, T., & Smith, A. (2014). *Value proposition design.* John Wiley & Sons.
- Rufus Pollock. (2011, March 31). *Building the Open Data Ecosystem.* <https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/>.
- Sambra, A. V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., ... Berners-Lee, T. (2016). Solid: a platform for decentralized social applications based on linked data. *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep.*
- Schweihoff, J., Jussen, I., Stachon, M., & Möller, F. (2022). Design options for data-driven business models in data-ecosystems.
- Sorescu, A. (2017). Data-driven business model innovation. *Journal of Product Innovation Management*, *34*(5), 691–696. DOI: 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). Retrieved from <https://doi.org/10.1016/j.giq.2022.101763> DOI: 10.1016/j.giq.2022.101763
- Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives.
- Van Lancker, D., Colpaert, P., Delva, H., Van de Vyvere, B., Meléndez, J. R., Dedecker, R., ... Verborgh, R. (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.
- 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. DOI: 10.1177/17835917221141324
- Wernick, A., Olk, C., & von Grafenstein, M. (2020). Defining data intermediaries. DOI: 10.26116/techreg.2020.007
- Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information polity*, 19(1-2), 17–33.