



# Establishing and governing data ecosystems at the crossroads of centralization and decentralization

Philipp Kernstock<sup>1</sup> · Constantin Harms<sup>1</sup> · Andreas Hein<sup>2</sup> · Helmut Krcmar<sup>1</sup>

Received: 22 March 2024 / Accepted: 11 June 2025  
© The Author(s) 2025

## Abstract

Data ecosystems are increasingly central to organizational strategy as they promise to democratize data sharing and enhance sustainability through collaborative models. Grounded in theories of decentralized governance, we examine how these ecosystems evolve from a conceptual decentralized framework to a more centralized operational reality as they mature. Employing an exploratory case study of four data ecosystems, based on 25 interviews and archival data, we investigate the transition within data ecosystems from decentralized emergence to the governance trade-offs necessitated by their expansion and increased complexity. Our findings depict a spectrum of governance adaptations: while some ecosystems develop formal structures that lean towards centralization to facilitate scaling, others maintain their foundational decentralized approach through self-regulation and technology-driven solutions. Our results contribute to the theoretical understanding of the dynamic governance within data ecosystems, revealing the processual nuances of balancing decentralization with operational centralization. This has implications for practitioners who must design flexible governance mechanisms capable of navigating between decentralized ideals and the centralizing demands of ecosystem growth and complexity.

**Keywords** Data ecosystem governance · Decentralized governance · Data ecosystem emergence · Case study research · Governance tensions

**JEL Classification** O32 · O33 · O39

---

Responsible Editor: Boris Otto

✉ Philipp Kernstock  
philipp.kernstock@tum.de

Constantin Harms  
c.harms@tum.de

Andreas Hein  
andreas.hein@unisg.ch

Helmut Krcmar  
helmut.krcmar@tum.de

<sup>1</sup> School of Computation, Information, and Technology, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

<sup>2</sup> Institute for Information Systems and Digital Business, University of St. Gallen (HSG), Müller-Friedberg-Strasse 8, 9000 St. Gallen, Switzerland

## Introduction

Organizations today face the challenge of addressing complex issues, such as meeting sustainability goals or improving operational efficiency. These efforts increasingly depend on data-driven collaboration among stakeholders, which requires a foundation of equality and mutual trust to ensure data is shared and utilized effectively to achieve collective objectives (Curry & Ojo, 2020; Otto et al., 2020). In light of these challenges, data ecosystems have emerged as a promising model to decentralize and democratize how data is shared among stakeholders (Azkan et al., 2022; Möller et al., 2024; Otto & Jarke, 2019). Data ecosystems provide a unified, secure, and trusted space to share, use, and consume data across stakeholders (Oliveira & Lóscio, 2018).

The practical relevance of data ecosystems is evident across diverse sectors. In the automotive industry, Catena-X exemplifies how data ecosystems can enhance product traceability throughout a vehicle's lifecycle. By creating a neutral, shared data space, Catena-X enables all actors

in the automotive value chain to contribute and access standardized lifecycle data, facilitating transparency in quality assurance, regulatory compliance, and carbon footprint assessments (Catena-X Automotive Network, 2025). Similarly, in the healthcare sector, the HEALTH-X dataLOFT initiative seeks to overcome longstanding data silos by enabling citizens to contribute their health data to scientific research and personalized health services. It adheres to Gaia-X principles, ensuring data sovereignty and user control despite cross-organizational data sharing (HEALTH-X dataLOFT, 2025). In mobility, the Mobility Data Space (MDS) facilitates decentralized data exchange through a federated infrastructure and data catalog. By connecting public and private data providers, MDS supports the development of services such as congestion reduction and mobility optimization (Mobility Data Space, 2025).

Despite these examples' promise, their implementation reveals a pressing governance challenge. As data ecosystems scale and diversify, the operational complexity of coordinating multiple actors, aligning data standards, and ensuring security and trust grows significantly. In response, many ecosystems have introduced new coordinating roles, such as app store providers, orchestrators, or operating companies that assume responsibilities for standard enforcement, certification, and quality assurance (Azkan et al., 2022; Otto & Jarke, 2019). While such roles enhance operational manageability, they also mark a structural departure from the original vision of decentralization. These shifts risk undermining the foundational principles of equal participation and mutual trust by concentrating control in the hands of coordinating entities. The resulting tension between the need for centralized orchestration and the aspiration for decentralized governance poses a critical and unresolved dilemma regarding how data ecosystems can sustain the ideals of decentralized data sharing while ensuring operational viability at scale.

Research on data ecosystems has concentrated on their initial development stages, highlighting the importance of decentralized models for data management across supply and value chains (Gelhaar & Otto, 2020; Otto & Jarke, 2019). This work examines various aspects, including role definition, trust-building, ensuring equal access to secure participant engagement, and establishing consensus on legal and governance frameworks (De Prieelle et al., 2022; Gelhaar & Otto, 2020; Gieß et al., 2023; Lefebvre et al., 2023). However, as these ecosystems mature, they face increasing complexity, often prompting a drift toward more centralized governance structures (Aaser et al., 2020). The implications of this drift for maintaining equal access and mutual trust and the strategies data ecosystems can employ to balance the demands of decentralization with the challenges of complexity and centralization are still unresolved (Möller et al., 2024; Otto & Jarke, 2019). This unresolved question marks a

significant gap in our understanding of the evolving dynamics within data ecosystems.

While data ecosystems are grounded in decentralization, which promises data sovereignty, fairness, and mutual trust (Otto & Jarke, 2019; Schleimer et al., 2023), the increasing complexity in practice often necessitates a drift toward more centralized data governance and control models, for example, by introducing operating companies. This shift from decentralized ideals to more hybrid real-world practices has important implications for the governance and evolution of data spaces. Although the literature provides insights into the structural and functional aspects of early-stage data ecosystems, the emerging centralization-decentralization tension of evolving ecosystems remains underexplored (Azkan et al., 2022; De Prieelle et al., 2022; Otto & Jarke, 2019). Our research closes this gap by investigating *how emerging data ecosystems overcome the challenges of decentralization by managing a centralization-decentralization trade-off*.

To address this gap, we employed an inductive multiple-case study, which is well-suited for exploring complex, emerging phenomena and building theory from rich empirical data (Eisenhardt, 1989; Eisenhardt & Graebner, 2007). Our sampling strategy focused on four data ecosystems (Auto-Eco, Mobility-Eco, Aero-Eco, and Marine-Eco) with varying degrees of centrality to observe their effects on governance evolution. This variation allowed us to capture divergent approaches to managing the balance between decentralization ideals and the practical necessities of centralization. We collected 25 semi-structured interviews with consortium members and archival records for triangulation and validity. Based on inductive coding (Strauss & Corbin, 1990), we derived insights on data ecosystem emergence drivers, value co-creation opportunities, development of standards, and operationalization tensions as factors influencing the trajectory of data ecosystems from growth to complexity. Following a cross-case analysis (Eisenhardt, 1989), we identified patterns in how these complexities led to strategic governance shifts.

Consequently, our research delineates two distinct phases of data ecosystem evolution: the decentralized emergence and the subsequent governance trade-off. The initial phase is hallmarked by ideation grounded in democratized and sovereign data sharing, where ecosystems ambitiously set out to develop a breadth of use cases underpinned by advanced standards. However, this pursuit of decentralized ideals is met with the complexities of orchestration, leading to what we term the “operationalization tension.” At this juncture, ecosystems confront the governance trade-off, navigating the delicate balance between the foundational decentralized principles and the centralizing forces that emerge as they scale. Our findings reveal divergent approaches to this evolution. Ecosystems with broader and more complex aspirations tend to form formal orchestration structures, shifting towards

organizational centralization for scaling. Conversely, those with a narrower focus use established standards and technical solutions to support a decentralized model reliant on self-regulation. These differences underscore the contingent nature of governance evolution in data ecosystems, where specific operational contexts and strategic objectives shape the extent and form of centralization.

From a theoretical perspective, the results contribute to understanding data ecosystems (Gelhaar et al., 2021a; Oliveira & Lóscio, 2018; Scheider et al., 2023) by delineating the processual dynamics of governance evolution within the context of centralization and decentralization tensions. The model presented shows the dynamic governance adaptation within data ecosystems, bridging the gap between decentralization ideals and the practical necessities of centralization for managing complexity and ensuring sustainable scalability (Beverungen et al., 2022; Chen et al., 2021). Practically, the study offers insights for practitioners involved in the design, management, and evolution of data ecosystems, suggesting that a flexible approach to governance that allows for shifts between decentralized and centralized models may be necessary to navigate the complexities of ecosystem growth.

## Theoretical background

We start the theoretical background by introducing the tenets of data ecosystems, their relation to business ecosystems, and decentralized governance. Next, we synthesize research on the emergence and evolution of data ecosystems to demonstrate the predominant focus on the early stages. Last, we conceptualize potential tensions between an initial, decentralized approach and a more centralized approach that develops due to the increasing complexity that data ecosystems face in later stages.

### Data ecosystems as digital ecosystems

Originating from biological systems of interacting organisms placed in a habitat, Moore (1993) introduced the concept of “ecosystems” to the business literature. Iansiti and Levien (2004) then used the concept to describe the mutual dependencies of loosely interconnected actors in business networks organized around keystone actors. This perspective focuses on actors being affiliated with an ecosystem. From a structural perspective, ecosystems can also be defined as “the alignment structure of the multilateral set of partners that need to interact in order for a focal value proposition to materialize” (Adner, 2017, p. 42). In innovation ecosystems, actors coordinate their non-generic complementarities to create interrelated innovations (Jacobides et al., 2024).

Extending the structural view of ecosystems to the digital realm, scholars have begun conceptualizing data ecosystems as configurations where data, not just products or services, becomes the core element of value creation and coordination (Beverungen et al., 2022; Möller et al., 2024). Data ecosystems are “a set of networks composed by autonomous actors that directly or indirectly consume, produce or provide data and other related resources (e.g., software, services, and infrastructure)” (Oliveira & Lóscio, 2018, p. 4). Akin to the structural perspective of ecosystems, they center on dynamic inter-organizational data sharing based on common value drivers instead of merely executing fundamental business functions through bilateral data sharing enabled by inter-organizational information systems (Möller et al., 2024). Closely related to the concept of data ecosystems is data spaces. Data spaces primarily focus on the technical infrastructure that facilitates data sharing within their ecosystem of autonomous actors (Gieß et al., 2023). Foundational work on data spaces has characterized them as a multi-sided data platform often driven by alliances (Otto & Jarke, 2019), and recent work has defined them as “decentralized data infrastructures designed to enable data-sharing scenarios across organizational boundaries by implementing mechanisms for secure and trustworthy data sharing – such as distributed data storage and the sharing of meta-data” (Möller et al., 2024, p. 6).

Recent research on data ecosystems has investigated challenges to their emergence (Gelhaar & Otto, 2020), relevant roles for value co-creation in data ecosystems (Azkan et al., 2022), the role of data governance in such ecosystems (De Prieëlle et al., 2022), and the design of B2C data ecosystems as opposed to industrial initiatives (Scheider et al., 2023). Further research has aimed at creating a shared understanding by developing taxonomies for data ecosystems (Gelhaar et al., 2021a), data sharing within them (Gelhaar et al., 2021b; Jussen et al., 2024), and their governance models (Kari et al., 2025; Lis & Otto, 2021). A particular focus is on the early stage and the implementation of a decentralized architecture (Gelhaar et al., 2021a) to ensure transparency and trust in data exchange between firms with intricate relationships characterized by simultaneous coopetition (Gelhaar & Otto, 2020; Hein et al., 2019). Hence, this research influences and is influenced by data ecosystem initiatives such as the automotive data ecosystem Catena-X or the mobility data marketplace MDS, which promote a decentralized ideal and aim for more equitable control and data sovereignty (Gelhaar et al., 2021a; Möller et al., 2024; Otto & Jarke, 2019). Despite the importance of decentralized governance models within these data ecosystem initiatives, how these so-called alliance-driven data ecosystems and the enabling digital infrastructure emerge and evolve is little researched and understood (Gelhaar et al., 2023; Möller et al., 2024; Otto & Jarke, 2019).

## Emergence and evolution of data ecosystems

Ecosystems do not emerge spontaneously but as the result of a process. Initiating firms craft rules and shape the process of ecosystem development to tie in complements and make complementors abide (Jacobides et al., 2018). However, as ecosystems emerge around a shared value proposition, their design can only be planned to a certain extent (Adner, 2017). Consequently, data ecosystems emerge through a complex orchestration process that has to overcome competitive and cooperative challenges (Gelhaar & Otto, 2020).

While the infrastructural foundation of a data ecosystem may be either centralized or decentralized (Gelhaar et al., 2021a), and both setups can lead to the successful implementation of data ecosystems (Kernstock et al., 2025), the prevailing trend is toward decentralized structures inspired by the ideals of the International Data Spaces (IDS). IDS provides a reference architecture, a formal standard, and reference implementations and aims to establish standards for trustworthy, sovereign, and self-determined data exchange. Rather than a technical platform, IDS is a concept based on peer-to-peer communication similar to the Internet (Lis & Otto, 2021; Otto & Jarke, 2019). This means authority and governance are not concentrated in a single locus within such ecosystems but are distributed across multiple control centers. The decentralized structure is pivotal for providing a unified, secure, and trusted space to share, use, and consume data across stakeholders (Oliveira et al., 2019). Equal access, data sovereignty, and mutual trust are crucial in the B2B settings where data ecosystems emerge, as joining industry ecosystems dominated by competitors is undesirable for most organizations (Abbas et al., 2024; Gelhaar & Otto, 2020). Thus, the decentralized conceptualization of data ecosystems addresses the issues resulting from ecosystem actors simultaneously engaging in competitive and cooperative relationships by ensuring equality and mutual trust (Adner, 2017).

Digital infrastructures like data spaces, which act as the foundation of data ecosystems, possess the evolutionary mechanisms of innovation, adoption, and scaling (Henfridsson & Bygstad, 2013). Otto and Jarke (2019) describe the evolutionary path of IDS as an alliance-driven multi-sided data platform leading from adoption to innovation to scaling. While the decentralized tenet might hold in the adoption phase through working groups, committees, and similar efforts (Kari et al., 2025; Otto & Jarke, 2019), relying on them to innovate and scale the platform might prove challenging as the decentralized structure leads to increasing complexity with a growing ecosystem due to the absence of an orchestrating entity (Jacobides et al., 2018; Parker et al., 2017). Practical examples like Catena-X and MDS show that increasingly complex data ecosystems shift towards establishing central institutions to govern the ecosystem's emergence and evolution. However, existing research on these

alliance-driven data ecosystems contends they have a decentralized, distributed organizational form aiming for fairer control and data sovereignty (Gelhaar et al., 2021a, 2023). Due to data ecosystems being a nascent practical phenomenon, research on their later evolutionary phases is scarce, and existing research does not consider how the tenets and roles of decentralized governance of data ecosystems might change over time (Brée et al., 2024; Oliveira et al., 2019).

## Centralization and decentralization in data ecosystems

The trade-offs between centralized and decentralized governance in digital ecosystems encompass complex tensions that influence their evolution and performance (Constantinides et al., 2018). Stability and evolvability are crucial requirements for technology ecosystems, highlighting the need for both homogeneity to benefit from standardization and heterogeneity to adapt to changing market demands. The inherent tensions concern the outputs (standard versus variety), actors (control versus autonomy), and identifications (collective versus individual) of and within an ecosystem (Wareham et al., 2014).

Centralized governance allows for decisively managing these tensions and implementing effective governance processes that enforce technical compatibility and quality control. Yet it may limit the ecosystem's adaptability and the variety of offerings (Chen et al., 2021; Hein et al., 2020). Further, without appropriate checks and balances, there is a risk of making decisions for the owners' interests rather than the ecosystem's collective good. Conversely, decentralized governance enhances participant power and inclusion in decision-making, aligning platform design with the interests of a broader community and potentially increasing the platform's welfare maximization (Chen et al., 2021; Ostrom, 2010). However, too much decentralization can be detrimental. When governance rights are excessively dispersed, the ability to mobilize collective action diminishes, and reaching consensus becomes challenging due to diverse participant perspectives. A semi-decentralized approach with key parties playing a role in governance can avoid such pitfalls, fostering effective governance while incorporating varied perspectives (Chen et al., 2021; Hsieh & Vergne, 2023).

In industrial data ecosystems, governance facilitates control and collaboration, which is essential for forging trust and enabling secure data exchange (Lis & Otto, 2021). Participating actors must independently establish governance mechanisms for their interactions. Thus, it is decentralized since the underlying infrastructure does not mandate how particular transactions are governed (Beverungen et al., 2022). Decision-making authority is dispersed among actors rather than being the sole purview of a focal entity (Lis & Otto, 2021). This open and inclusive approach to governance facilitates

the integration of varied perspectives. Addressing the challenges of consensus-building around standards and legal measures is crucial, as a lack of guidelines or cultural disparities among actors can often hinder this process (Fassnacht et al., 2023; Gelhaar & Otto, 2020). Hence, decentralized governance, supported by trust, decentralized architecture, and data exchange transparency, is vital for the health and progress of data ecosystems (Chakrabarti et al., 2018).

While the inherent tensions between centralized and decentralized governance in digital ecosystems are acknowledged in the literature, the mechanisms by which data ecosystems reconcile these tensions during their formative stages and form functioning governance modes for their expansion remain ambiguous. Literature primarily focuses on the emergent phases of data ecosystems (Gelhaar & Otto, 2020; Otto & Jarke, 2019), where governance structures are not yet crystallized, foundational policies and norms are being negotiated, and the ecosystem's trajectory is still malleable (Tilson et al., 2010). However, as examples such as Catena-X illustrate, data ecosystems increased in complexity over time and started incorporating centralized and decentralized governance mechanisms. How well these decentralized governance approaches can deal with the increasing complexity of later stages of data ecosystems while retaining their conceptual ideals remains to be understood (Beverungen et al., 2022). While recent research has shown that the governance structures of data ecosystems evolve (Kari et al., 2025), this shift and the nuances of this centralization-decentralization tension remain underexplored and inadequately understood from an academic perspective and present a vital research gap for understanding the later stages of data ecosystem development (Oliveira et al., 2019).

## Research design

We employed an explorative multiple-case study approach to investigate how emerging data ecosystems manage the centralization-decentralization trade-off (Eisenhardt, 1989; Eisenhardt & Graebner, 2007). This method is particularly suitable for examining complex, dynamically evolving phenomena within their real-life contexts (Siggelkow, 2007). Given the emergent nature of data ecosystems and the identified theory gap, an explorative, inductive multiple-case study enables us to generate new theoretical insights grounded in empirical data, which is essential for building a nuanced understanding of how varying degrees of centrality impact governance evolution in data ecosystems. In addition, a multiple case study allows us to compare and contrast different ecosystems, enhancing the robustness and generalizability of our findings (Yin, 2018). Following a within and cross-case analysis, we can identify patterns and variations across different settings, thereby strengthening theory development by

uncovering underlying mechanisms and contextual factors (Eisenhardt & Graebner, 2007).

## Data sampling

Our sampling strategy focused on selecting data ecosystems with varying degrees of centrality to observe their effects on governance evolution. To achieve this, we applied three sampling criteria for our multiple-case study (Eisenhardt, 1989; Eisenhardt & Graebner, 2007):

First, we selected data ecosystems based on three criteria: they were established by multiple organizations, had significant resource investments (multi-million Euro investments), and were either in the operationalization phase or already live and scaling. These criteria ensured that the chosen ecosystems had reached a level of maturity sufficient to provide substantial insights. Ecosystems at the operational or scaling stage are more likely to have encountered and addressed the challenges associated with transitioning from decentralization to centralization. Understanding their different governance approaches to managing these challenges is crucial for examining the trade-offs between these strategies. By focusing on ecosystems with significant investments, we ensured that any challenges identified were not merely due to insufficient funding.

Second, we sampled data ecosystems based on their degree of transition from decentralized ideals to centralized realities. Including ecosystems with varying approaches along the decentralization-centralization spectrum provides essential variation for theory-building (Eisenhardt & Graebner, 2007). This diversity allowed us to identify patterns, similarities, and differences in how different approaches address the centralization-decentralization tension. We began with Auto-Eco,<sup>1</sup> a purposeful sample (Siggelkow, 2007) representing the most prominent and well-funded data ecosystem initiative currently operational, based on a strong alliance of different automotive industry companies. Next, we selected Mobility-Eco, another prominent example centered around one orchestrating company. Both cases revealed a spectrum of approaches to building data ecosystems, ranging from complete decentralization to greater reliance on a keystone actor. It is important to note that, due to the general conception of data ecosystems and their underlying infrastructure, i.e., data spaces (Möller et al., 2024), even the "more centralized" approaches in our sample are far from being fully centralized as in traditional platform ecosystems; all maintain values of trust and data sovereignty. We then extended our case sample through Marine-Eco, which had a setup between the first two cases on the spectrum, and

<sup>1</sup> Case names are pseudonyms to maintain the anonymity of cases and interview partners.

Aero-Eco, which relied even more on decentralization, with infrastructure components based on blockchain technology.

Third, we selected only cases originating in Germany to account for national and regulatory differences. By focusing on a single country, we controlled for variations in national regulatory environments that could otherwise confound our findings. This focus allowed us to isolate the impact of the ecosystems' internal governance approaches rather than external regulatory factors. Through these criteria, we aimed to capture nuances in how different data ecosystems deal with the centralization-decentralization challenge. Table 1 gives an overview of the sampled cases.

All selected cases started as German data ecosystem initiatives, initially involving 10 to 28 participating firms, and were actively under development at the time of writing. The ecosystems were in the automotive, mobility, aerospace, and maritime sectors. Since all cases emphasized their orientation towards productive operation and came from different industries, they fulfilled our requirements. While all selected initiatives adhered to decentralized ideals such as data sovereignty, their approaches to the challenges of ecosystem emergence and evolution lay on a spectrum concerning the degree of centralization (see Fig. 1).

## Data collection

We collected data through 25 semi-structured interviews across four data ecosystems (Yin, 2018) between July 2023 and August 2024. The interviews aimed to capture the factors driving ecosystem development, the strategic responses to emerging complexities, and the governance choices made by firms in navigating between centralized and decentralized models. We developed the questionnaire for our semi-structured interviews based on existing guidelines for collecting qualitative data (Klein & Myers, 1999; Myers & Newman, 2007). Drawing on seminal literature on digital ecosystems by Adner (2017) and Jacobides et al. (2018), we designed our questionnaire to center around four key themes: (1) ecosystem coordination, which includes questions on how different actors join and align within the ecosystem; (2) ecosystem collaboration, which includes questions on what roles actors play and how they interact with each other; (3) ecosystem value creation and capture, which includes questions of how value is created and appropriated by the different actors; (4) ecosystem governance, which includes questions on how the overall development and evolution of the ecosystem is steered.

We strived to let our interview partners describe their world in their own words and varied the exact questions and the focus of the interviews based on the interviewee's background (Myers & Newman, 2007), e.g., when interviewing with technical roles, we focused our questions on technical aspects of governance (see Appendix 1). Table 2 provides an overview of the interviewees. All interview partners were

**Table 1** Case sampling overview

Case	Auto-Eco	Mobility-Eco	Aero-Eco	Marine-Eco
Industry	Automotive	Urban mobility	Aerospace	Maritime
Vision	Facilitate secure, standardized, and cross-company data exchange to enhance efficiency, sustainability, and innovation	Providing a data marketplace where equal partners can exchange data to make mobility safer, more efficient, and more environmentally friendly	Accelerating digital transformation of the aerospace industry, focusing on developing efficient, decentralized working methods and processes across the entire lifecycle of space and air vehicles	Creating a digital maritime data space based on data sovereignty, security, interoperability, and modularity
Participants	Automotive manufacturers, suppliers, (information) technology providers	Mobility service providers, authorities, public transportation companies, infrastructure providers	Aerospace manufacturers, suppliers, consulting companies	Software providers, infrastructure companies, research institutes
Number of use cases	11	9	8	5



**Fig. 1** Sampling strategy (numbers indicate sampling order)

**Table 2** Conducted interviews

Data ecosystem	ID	Line of business	Position	Length
Auto-Eco	AU1	Operating company	Product Lead	33 min
	AU2	Research institute	Department Head Logistics	49 min
	AU3	Multinational software provider A	Head of Industry Business Networks	50 min
	AU4	Multinational technology platform provider	General Manager Digital Sovereignty	31 min
	AU5	Automotive OEM A	Product Owner	39 min
	AU6	IT service provider	Business Executive	46 min
	AU7	IT consultancy	Principal Project Manager	38 min
	AU8	Multinational engineering company	Project Director	46 min
	AU9	Multinational software provider A	Principal Architect	47 min
	AU10	Automotive OEM B	Head of Digital Production	43 min
	AU11	Automotive supplier	Head of Data Space Development	35 min
Mobility-Eco	MO1	Operating company	Technical Expert	49 min
	MO2	IT startup	Business Development	21 min
	MO3	IT startup	CEO	42 min
	MO4	German public transport company	Team Lead	37 min
	MO5	Operating company	Managing Director	48 min
Marine-Eco	MA1	German internet service provider	SVP Digital Ecosystems	38 min
	MA2	German internet service provider	Senior Lead Researcher	45 min
Aero-Eco	AE1	German aerospace research institute	Project Manager	32 min
	AE2	Large system integrator	Digital Transformation Manager	22 min
	AE3	Aerospace service provider	Project Manager	28 min
	AE4	IT startup	CEO	28 min
	AE5	German research institute	Research Fellow	35 min
	AE6	German aerospace supplier	Project Engineer	33 min
	AE7	IT startup	Software Engineer	44 min

part of companies that have participated in their respective data ecosystem initiatives since their inception. Consequently, they were able to offer insights into the journey from the ecosystems' emergence to their subsequent evolution. Notably, the interviewees who worked for the later emerging operating companies were involved in the respective ecosystem initiatives in the earlier stages as employees of ecosystem member firms.

Archival sources, including public presentations, whitepapers, and technical documentation, supplemented our data, providing a richer context for analyzing the strategic maneuvers within these ecosystems (see Appendix 2). We selected sources with direct pertinence to the cases under study as they offer contextual depth, allowing us

to triangulate the insights extracted from the interviews. Further, there is a lack of independent reports, making the official documents essential to gather sufficient data. This approach allowed for an in-depth examination of how firms adjust their governance structures in response to the challenges posed by evolving ecosystem dynamics.

### Data coding and analysis

The iterations between data collection, within and cross-case analysis, and comparison with existing literature guided our research (Gioia et al., 2013). We coded the interviews according to 1st-order concepts, 2nd-order themes, and aggregate dimensions using MAXQDA. Following this

inductive coding process (Gioia et al., 2013; Strauss & Corbin, 1990), our analysis delineated the evolution within data ecosystems, focusing on the nexus between ecosystem expansion, resultant complexities, emerging tensions, and the strategic drift towards centralized or decentralized governance mechanisms. To ensure analytical coherence and allow for deep immersion in the data, the first author conducted the initial round of coding. Rather than conducting independent coding, we prioritized a dialogic process, where the first author extensively discussed the coding structure, interpretations, and emergent patterns with the second author. This iterative exchange helped critically challenge and validate the initial codes, fostering a shared understanding. Throughout each stage of the aggregation process, the authors debated the relevance and accuracy of each code. They reflected on the emerging themes, considering how best to represent the nuances of the data. This collaborative and iterative approach led to adjustments in the coding structure, ensuring it was comprehensive and coherent. This approach aligns with established practices in inductive research, including the Gioia methodology, where one researcher often performs initial coding before collaborative refinement, avoiding the risk of “going native,” i.e., being too close to the case and adopting the informants’ perspectives (Gioia et al., 2013). The constant comparison with the cases allowed the authors to fine-tune the codes, making them more precise and aligned with the data (Strauss & Corbin, 1990). The collected secondary data was used to validate and triangulate the insights from the interviews and the developed codes (Yin, 2018).

Following this process, we developed and defined the six aggregate dimensions (as depicted in Fig. 2). To verify the robustness of our coding structure, the third author, uninvolved in initial coding, deductively applied the aggregate dimensions. This step served to validate the coding framework from an independent perspective. The coding was discussed in case of ambiguity or differences to agree on coding decisions and clarify definitions.

In the initial phase of our analysis, we conducted within-case inductive coding, revealing that data ecosystems emerged in a decentralized manner driven by specific factors. We identified open codes such as “more fairness in the data economy” and “less centralized power of infrastructure providers,” which illustrate the political ideals motivating the emergence of these ecosystems. We combined these codes under the axial code “political factors.” Additionally, we found open codes like “higher resilience through tightening of whole industries,” meaning that data exchange strengthens cooperation across entire industrial sectors beyond competitive relationships. These were grouped under the axial code “ecosystem thinking.” As these ecosystems evolved, we observed that increasing complexity led to operational tensions. We identified open codes such

as “extent of governance participation” and “enablement of integration and connection to ecosystem,” illustrating the direct outcomes of ecosystem expansion. We grouped these codes under axial codes like “operational complexity” and “governance fragmentation,” capturing the broader themes of increasing intricacies and strain on existing governance structures. Finally, we combined both axial codes under the selective code “operationalization tension,” which addresses the issues faced when scaling data ecosystems.

Building upon our initial findings, we conducted a cross-case analysis to uncover variations in how each data ecosystem emerged and the extent to which they shifted away from their decentralized ideals to manage growing complexities. This analysis revealed patterns of similarities and differences in how the evolution of these ecosystems led to varying degrees of governance fragmentation and operational complexity, focusing on the four aggregate dimensions developed in the decentralized emergence phase (Yin, 2018). For instance, in the Auto-Eco case, we observed that the ecosystem grappled with integrating organizations of varying sizes necessary to realize its use cases. Committed to its democratic ideals, Auto-Eco made its core components open source. However, this approach resulted in difficulties ensuring continuous and equitable financing for the further development of these open-source components. Conversely, Aero-Eco, being a smaller data ecosystem, had to navigate the stringent regulations of the aerospace sector. It encountered obstacles with the Gaia-X framework it relied on, ultimately leading it to adopt a new technical infrastructure. Ultimately, we observed a common struggle across all initiatives to transform their decentralized ideals into operational reality.

As we advanced to the second phase of our analysis, we focused our inductive coding on how previously identified operational tensions led ecosystems to engage in governance trade-offs. The within-case analysis uncovered open codes such as “establishment of operating units” and “division of powers and control,” highlighting the divergent strategies ecosystems employed in response to the complexities they faced. We refined these into axial codes like “operating model development” and “governance mechanism establishment,” depicting ecosystems’ dual paths to address their internal challenges. The selective code “data ecosystem operationalization” captures these diverging efforts to scale the ecosystems. This coding phase illustrates a critical juncture where ecosystems, confronted with the tensions brought about by increasing complexity, made governance trade-offs. Some shifted towards more centralized governance to enhance efficiency and control, while others maintained their decentralized ethos to preserve autonomy and inclusiveness.

Building on these insights, we conducted a cross-case analysis of the four data ecosystem initiatives using our established coding framework (see Appendix A4). This

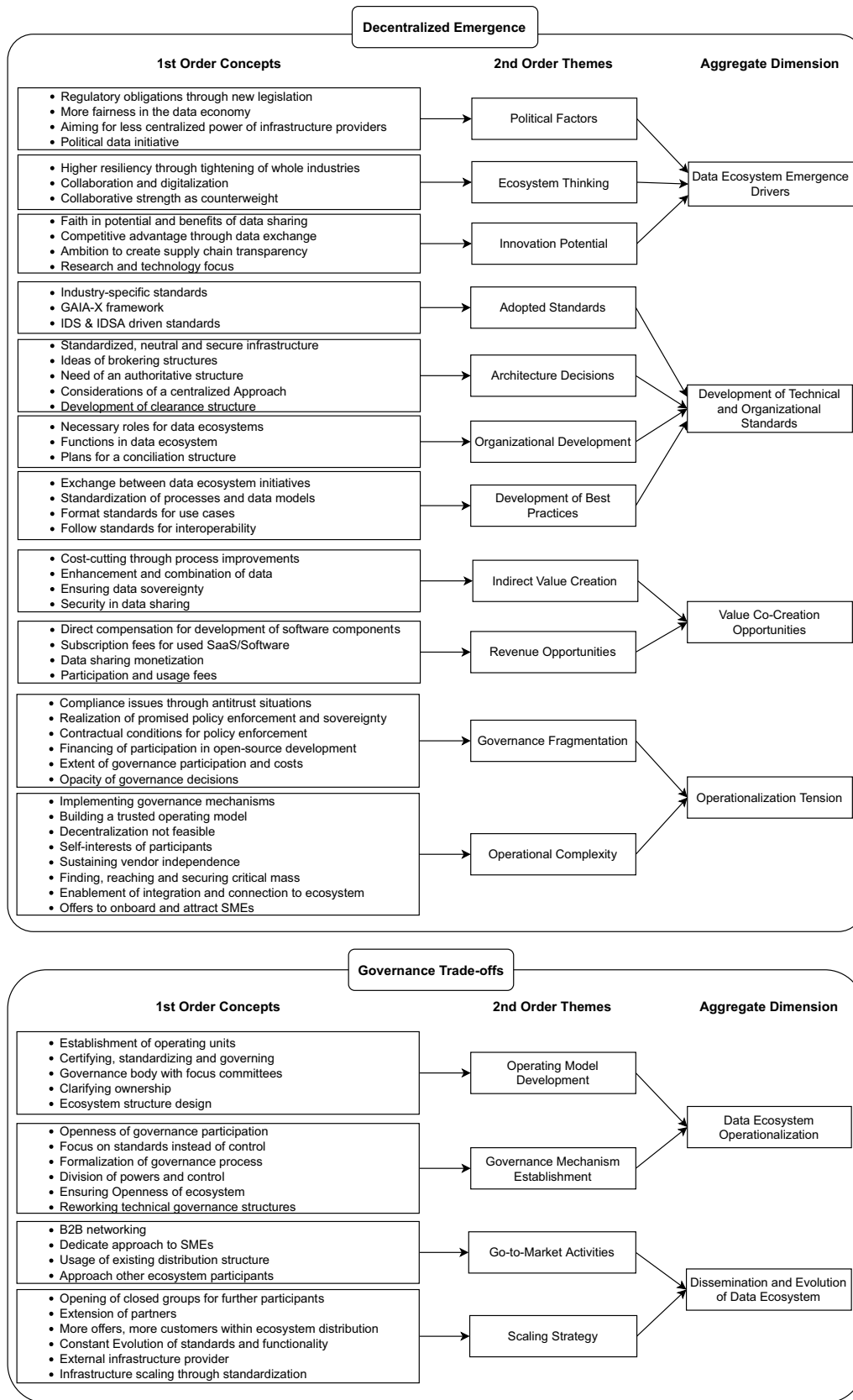


Fig. 2 Overview of the inductive data structure (based on Gioia et al., 2013)

analysis focused on the two aggregate dimensions identified in the governance trade-offs phase to identify variations in the ecosystems' responses to organizational tensions. While all ecosystems grappled with the tension between centralization and decentralization, they varied significantly in their emphasis on evolutionary mechanisms such as adoption, innovation, and scaling (Henfridsson & Bygstad, 2013). This step enhanced the robustness of our findings by allowing us to systematically cluster the ecosystems based on their responses to organizational complexity and governance fragmentation.

## Results

The within-case analysis revealed six aggregate dimensions, which we structure based on the decentralized emergence and governance trade-off phases. The cross-case analysis illustrates how the four data ecosystems addressed the four dimensions of the decentralized emergence phase and the two dimensions focusing on managing the trade-off between centralization and decentralization. While there were commonalities and differences among all cases, the Auto-Eco and Mobility-Eco cases were similar in that they built large-scale data ecosystems to enable data sharing and usage across large industrial sectors, and that this scale led to a stronger need to mitigate the operationalization tensions through governance trade-offs. Auto-Eco and Mobility-Eco prioritized

innovation to navigate complexity, leading them to shift towards centralized governance structures to streamline decision-making and accelerate the development of new solutions. These initiatives aimed to establish a de facto standard data ecosystem for the entire automotive and mobility sectors. The Aero-Eco and Marine-Eco cases addressed a smaller set of actors while operating in highly specialized industries, i.e., the aerospace industry being highly regulated and the maritime sector having highly specialized requirements hardly transferable to other industries. Aero-Eco and Marine-Eco emphasized a decentralized approach to preserve autonomy and foster a participatory environment conducive to scaling. They chose to adopt existing developments, such as data-sharing standards or technological infrastructure, to address challenges in their specialized environments.

We use the insights of the within and cross-case analysis to structure the results along the two types of ecosystem initiatives, i.e., large-scale and specialized, and along the aggregate dimensions developed through the inductive coding and structured by the decentralized emergence and governance trade-off phases (see Table 3). Since the large-scale and specialized data ecosystems shared similar patterns regarding their decentralized emergence but differ significantly in how they operationalized their ecosystems by engaging in governance trade-offs, we structure the results accordingly. First, we describe the shared patterns of decentralized emergence. Second, we detail the divergent patterns of governance trade-off.

**Table 3** Summary of case study analysis divided by development phases and ecosystem clusters

Phase		Large-scale data ecosystems (Auto-Eco, Mobility-Eco)	Specialized data ecosystems (Aero-Eco, Marine-Eco)
Decentralized emergence	<i>Data ecosystem Emergence drivers</i>	Emergence from a decentralized ideal of democratized and sovereign data ecosystems	Adoption and usage of the developed standards
	<i>Development of technical and organizational standards</i>	Development of advanced technical and organizational standards	Focus on a select number of use cases
	<i>Value co-creation opportunities</i>	Development of large number of use cases	Initial progress hits roadblocks through context-specific issues not addressed by existing standards leading to operationalization tensions
	<i>Operationalization tension</i>	Establishment of mutual trust and data sovereignty Growing complexity and orchestration effort leading to operationalization tensions	
Governance trade-offs	<i>Data Ecosystem Operationalization</i>	Establishment of operating entities for orchestration Larger drift toward organizational centralization to enable scaling	Addressing challenges through technology Smaller drift due to reliance on self-regulation and informal structures
	<i>Dissemination and Evolution of Data Ecosystem</i>	Developing scaling strategy and launching go-to-market activities	Realizing a select number of use cases with reduced scale
	<i>Managing Trade-off between Centralization and Decentralization</i>	Managing trade-off by balancing organizational centralization and technical decentralization	Remaining closer to initial concept by focusing on technical decentralization Less ecosystem scaling

## Shared patterns of decentralized emergence

The decentralized emergence of the data ecosystems was characterized by adoption and innovation activities. In this phase, the conceptual design was pushed by emergence drivers that led to the setup of data ecosystem initiatives and propelled a decentralized ideal. All ecosystem cases were focused on decentralization in their technical realization. Security, privacy, compliance, and freedom of choice were crucial for all initiatives. Participants should all have sovereignty over their data and decide whom to share it with and under which conditions. These rights were enforced contractually in the Auto-Eco, Mobility-Eco, and Marine-Eco, ultimately relying on trust between the parties. Aero-Eco went one step further, placing its infrastructure on top of blockchain technology and partially starting with technical enforcement via smart contracts. With that choice, Aero-Eco paved the way for complete decentralization. At best, Auto-Eco aimed to eliminate any remaining and arising centralities in its ecosystem over time. While this decentralized ideal fostered mutual trust of cooperative ecosystem actors, e.g., different automotive suppliers participating in Auto-Eco, and led to the establishment of data sovereignty, operationalization tensions arose later on as a consequence of the increasing complexity, the need to mediate decentralized ideals and the operational necessity of partial centralization.

### Data ecosystem emergence drivers

The data ecosystems emerged through a set of emergence drivers and motivational reasons of participants. The participating actors collaborated and formed initiatives, driving the data ecosystems' evolution. Specific drivers were already based on a decentralized ideal, while others led to its constitution through the necessary decisions to build a thriving data ecosystem. The data ecosystem initiatives initially adopted the developed decentralized ideal and drove their subsequent innovation activities.

Both large-scale and specialized data ecosystems were significantly shaped by **political factors**, **ecosystem thinking** and **innovation potential**. Large-scale ecosystems like Auto-Eco and Mobility-Eco were driven by regulatory mandates, particularly in the context of complex supply chains. The focus was on creating transparency and meeting compliance standards, with Auto-Eco participants expressing concerns about balancing openness with competition-related risks.

It's about how to open the supply chains without opening competition-related issues. (AU1 – Product Lead, Operating Company)

Specialized ecosystems, such as Aero-Eco, also responded to regulatory drivers, emphasizing preventing monopolization and fostering collaboration in highly regulated sectors. Pursuing a decentralized governance model was fueled by a collective desire for a less centralized power structure, offering a fairer environment where neutral entities, rather than dominant industry players, lead the ecosystems.

They wanted it to be managed by a more neutral entity than, for example, [a hyperscaler]. Yes, the idea behind the whole Gaia-X concept is also a bit of democratization to prevent any monopolization. (AE1 – Project Manager, Aerospace Research Institute)

**Ecosystem thinking** reflected a collective mindset among potential participants, which positively influenced collaboration and was a driving factor behind the emergence of both large-scale and specialized data ecosystems. This mindset was particularly focused on increasing the resilience of supply chains by tightening collaboration across industries. Events like pandemics and natural disasters highlighted the need for data access from distributed sources to effectively manage disruptions. However, this collaborative ecosystem approach sharply contrasted with the traditional relationships in industrial sectors, historically dominated by asymmetric power dynamics, for example, between aerospace or automotive OEMs and their suppliers.

After nobody was able to solve the problems for themselves [...] a concerted action became apparent [...] (AU5 – Product Owner, Automotive OEM)

**Innovation potential** was also a shared driver, though the scale and scope of these opportunities varied. Large-scale ecosystems focused on creating industry-wide business models and services, whereas specialized ecosystems concentrated on niche innovations. Access to diverse data sources that were not typically combined was seen as offering competitive advantages, protecting businesses from competitors and securing market positions.

[...] if I have any business ideas, any use cases, I want to implement, that I quickly get access to these [necessary] data (MA1 – SVP Digital Ecosystems, Internet Service Provider)

Thus, while both types of ecosystems offered significant innovation potential, large-scale ecosystems aimed for broader impact, and specialized ecosystems targeted specific regulatory and industry needs. Decentralization was critical in actualizing these perceived benefits and striving for a fair and collaborative environment with high transparency.

## Development of technical and organizational standards

Technical and organizational **standards** were critical for both large-scale and specialized ecosystems, providing the foundation for governance, operational protocols, and infrastructure development. Large-scale ecosystems, particularly Auto-Eco, played an active role in shaping these standards, contributing to frameworks like the International Data Space Association (IDSA). The adoption of standards has been driven by the formation of reference implementations and standards in recent years that guide data ecosystems, particularly highlighting the technical aspects of realization. Adopting frameworks from the International Data Space Association (IDSA) and Gaia-X stood out due to their mature and open nature, providing a foundation for decentralized governance.

The IDS is a good foundation of what is needed to be accomplished. So, IDSA was naturally quite theoretical but was a good fit from a theoretical perspective, and it was open and standardized. (AU4 – General Manager, Technology Platform Provider)

The specialized ecosystems like Aero-Eco and Marine-Eco primarily adopted these standards without playing a significant role in their development, instead tailoring them to meet their specific industry needs.

We are familiar with IDSA, the International Data Space Association, i.e. data spaces and data space standards, and we are cloud infrastructure providers. (MA1 – SVP Digital Ecosystems, Internet Service Provider)

**Architecture decisions** refer to the architecture of a data ecosystem, which serves as a technical core and is influenced by the adopted standards. Yet it may diverge to accommodate specific needs. Security, neutrality, and reliability surfaced as critical attributes for later development, with neutrality being vital for decentralized operations and governance.

But you need an infrastructure, which is secure, neutral, [...] and it is also obvious that you need structures which are industrial grade and reliable. (MO3 – CEO, IT Startup)

Large-scale ecosystems required robust, neutral infrastructures to manage complex operations, while specialized ecosystems like Aero-Eco pushed the boundaries of decentralization by implementing blockchain technology to manage transaction authorization. The cases followed a decentralization approach and thus needed a brokering structure that enabled the discovery of all the distributed data offers to allow participants to successfully use the data ecosystem. This, however, posed the risk of introducing some centralization as orchestration is needed. Further, the introduction

of centralized elements, in the form of authoritative structures for participant authentication and transaction authorization, highlighted a governance tension where the need for simplicity may inadvertently steer ecosystems towards centralization.

That is also a service, but that controls who is a participant in a federation and who is not. So, some authority, deciding and controlling how the federation is composed. And that is the central within the federation, the central which governs and controls. (AE5 – Research Fellow, Research Institute)

**Organizational development** activities created the organizational fabric of data ecosystems, which is defined by the roles and functions of the actors within and their mutual alignment, devoid of a definitive leader or governance model, focusing instead on self-regulation. The necessity for roles such as data providers, data consumers, and service providers emerged from use cases, with each role bearing distinct responsibilities.

There are always use cases which strongly determine which roles are necessary per use case. (MA1 – SVP Digital Ecosystems, Internet Service Provider)

Large-scale ecosystems, like Auto-Eco and Mobility-Eco, had to manage a more complex network of stakeholders, while specialized ecosystems were simpler and relied more on self-regulation. However, both recognized the potential need for formalized governance structures as the ecosystems grew, highlighting the tension between decentralized governance models and the practical need for some level of centralization.

**Best practices** that enhance the collective understanding of functioning within data ecosystems were developed as ecosystems evolved. These practices, born out of real-world implementation and iterative learning, fed back into the standards, enriching them.

We need to make sure that we standardize as many data models, as many processes, that we ensure this certainty [...], this reliability, this usage, that it really reaches an official state. (AU5 – Product Owner, Automotive OEM)

In large-scale ecosystems, best practices evolved to ensure operational consistency across a wide range of participants. In specialized ecosystems, best practices focused on addressing industry-specific challenges, like regulatory compliance in aerospace or maritime data management. It also highlighted the strategic importance of early consensus on architecture, which set a foundation for future interoperability and alignment with the ecosystem's decentralized governance principles. However, enforcing these developed best practices in a decentralized setup became a frequently mentioned issue.

Technical and organizational standards were pivotal to these ecosystems, acting as the backbone for integration and interoperability. The developed and adopted standards exemplified the push towards open, neutral, and reliable frameworks that are essential for decentralized governance. The decentralized ethos was evident in the preference for neutrality in infrastructure and operation. Roles and responsibilities emerged organically from use cases without a predefined leader or governance model. While this self-alignment was crucial to the ecosystem's functioning, it introduced complexity that necessitated the later formalization of governance mechanisms.

### Value co-creation opportunities

Data ecosystems must demonstrate their ability to forge clear value-creation paths for all stakeholders.

**Indirect value creation** opportunities made up a significant share of the use cases in the emergent phases of data ecosystems. Use cases often focused on deriving benefits through cost reduction and process enhancements enabled by shared data. Large-scale ecosystems like Auto-Eco and Mobility-Eco aimed to eliminate redundant processes and bureaucracy by shifting from manual to digital, data-driven systems.

In the supply chain context, there is a high portion of analog processes, manual tasks especially concerning information exchange [...] (AU5 – Product Owner, Automotive OEM)

Specialized ecosystems, like Aero-Eco and Marine-Eco, focused on improving process efficiencies in highly regulated sectors, such as ensuring compliance or optimizing data processing.

By collaborating, data ecosystem participants could enhance and combine data, thus increasing its value. For collaboration to be consensual and successful, all participants had to be clear about which data is exchanged, how exactly, and with whom, and feel comfortable with the overall transaction. Consequently, data sovereignty and security in data exchange played crucial roles in all value-creating activities.

[...] every participant in data sharing and the ecosystem should feel certain that their data is always sovereign and having full control over how the data is exchanged. (MO1 – Technical Expert, Operating Company)

This transformation was elemental to the decentralized ethos of the data ecosystem initiatives, emphasizing the collective benefit and shared value creation over individual gains. Realizing this freedom and preventing the manifestation of self-interest, decentralized architectures were favored by ecosystem participants, as they formed the basis for value co-creation and acted as a counterbalance to one-sided power plays.

New **revenue opportunities** varied between large-scale and specialized ecosystems. Large-scale ecosystems offered a broader range of revenue models, including subscription-based models, compensation for software development, and peer-to-peer transactions for data exchange. In contrast, specialized ecosystems like Aero-Eco and Marine-Eco focused on niche revenue opportunities tied to specific industry needs. However, incentivizing data sharing remained a debated topic within the ecosystem, as it touched upon the governance tensions between decentralized autonomy and the need for centralized mechanisms to ensure fair compensation and economic viability.

When somebody offers their data, they are also incentivized in return. Meaning when I give you data, I get money from you. (AU3 – Head of Industry Business Networks, Software Provider)

Operating and developing an ecosystem was not an end in itself for the participating actors. Its continuance and self-sufficiency must be at least secured. Subscription and participation fees provided a business model opportunity for operators and developers. The decentralized governance model allowed participants to form peer-to-peer agreements for use cases without reliance on a central authority. Such models reflected the ecosystems' aim to balance collaborative value creation with individual actor empowerment, navigating the inherent tensions between centralized control and decentralized innovation.

### Operationalization tension

The decentralized emergence led to mutual trust within the data ecosystems based on the actors retaining data sovereignty. However, they also increased complexity as the ecosystems evolved and grew. Thus, as data ecosystems transitioned from conceptual frameworks to operational entities, the gap between strategic ideals and the realities of implementation, an operationalization tension, became evident.

**Governance fragmentation** describes the issues resulting from increasing expansion through joining actors. The initial open invitation to a diverse array of firms, including competitors, set the stage for complex participatory challenges that drove a fragmentation of governance structures. In ecosystems like Auto-Eco, where major industry players converged, antitrust considerations and the exchange of sensitive information necessitated a robust framework for compliance and trust. Whenever data was exchanged, trust in enforcing regulations and data sovereignty was required.

When I now share my data, I might share it with you. But can I be certain that you do not forward it? (AU2 – Department Head Logistics, Research Institute)

Contracts became pivotal for ensuring data usage rights, yet the essence of trust in decentralized ecosystems hinged

on the belief in mutual adherence to these agreements. Technical solutions, such as those inspired by blockchain-based smart contracts implemented by Aero-Eco, sought to enforce these contracts programmatically. However, the reliance on open-source contributions and the inherent participation costs raised questions about power dynamics and decision-making transparency within the ecosystem. A few companies dominated governance and development decisions, creating a big player circle within the broader participant community.

**Operational complexity** for ecosystem initiators resulted from manifold challenges concerning the operationalization of the initiatives. Establishing governance mechanisms that balance decision-making authority and control was crucial yet nontrivial. The debates around the operator model, whether to adopt an independent operator to avoid bias and promote inclusivity, reflected the tensions between centralization and the decentralized ethos.

Who is operating the data space? Ideally somebody independent else half the participants are excluded because they do not want to join. (MA1 – SVP Digital Ecosystems, Internet Service Provider)

Trust among competing participants, effective policy enforcement, and the realization of data sovereignty were foundational to the ecosystems' success. Yet these goals were challenged by the personal interests that surfaced when formal governance structures were introduced.

[...] the moment you have control committees, the moment you have presiding structures, that moment it gets instantly personal. (AU4 – General Manager, Technology Platform Provider)

The aspiration for a fully decentralized ecosystem often clashed with the current organizational and technical readiness states, revealing a tendency to start with centralized architectures. While data ecosystems aimed to evolve into decentralized platforms, the transition from the centralized starting point to a decentralized end state proved complex. Thus, the quest for decentralization was often at odds with immediate operational needs and the vested interests of participants.

We cannot build this completely decentralized from one day to the next because the customers are not ready yet. (AU1 – Product Lead, Operating Company)

Centralized entities, such as the operating companies in the Mobility-Eco ecosystem or the Auto-Eco ecosystem, emerged as intermediaries to facilitate governance and orchestration. This revealed the practical push towards centralization that contrasts with the decentralized governance model advocated by the conceptual frameworks. While large-scale and specialized ecosystems shared these patterns,

the scale, complexity, and industry-specific focus of each type influenced how these dynamics played out. The tension between decentralization ideals and practical governance and operational needs was a recurring theme across both types, setting the stage for the governance trade-offs in the subsequent phase of ecosystem evolution.

### Diverging patterns of governance trade-offs

The data ecosystems initiatives engaged in a governance trade-off to address the arising complexities that led to operationalization tensions. Organizational approaches diverged notably across the cases studied. Governance was approached distinctively by the large-scale and specialized data ecosystems despite all encouraging open ecosystem participation by being receptive to new members. Characterized by scaling and continued innovation activities, more substantial governance trade-offs in the form of centralized elements became apparent during the operationalization of the large-scale data ecosystems. The conceptual design of the operational data ecosystems addressed the complexities materializing in the emergence phase but introduced tensions with the decentralized ideal.

#### Data ecosystem operationalization of large-scale data ecosystems

In its emergence, Mobility-Eco leaned towards a more centralized structure, while Auto-Eco favored a decentralized model with a more even power distribution captured in a newly established association. Nonetheless, both underscored the importance of transparency, employing certification, formalized rules, and controls. The shift towards production and the growing complexities led to the setup of a centralized entity within Auto-Eco, moderating the initially strongly decentralized ideal by establishing a federated form of governance, which relied on decentralized and centralized aspects. Already firmly in their scaling process, both mature cases were equipped with federated governance approaches where decentralized approaches were mediated with established organizational approaches.

**Developing an operating model** was crucial for the data ecosystems' further development and scaling, as it detailed the technical operation, governance structure, ownership, and overarching design. Mobility-Eco and Auto-Eco emerged from publicly funded projects, where the founding actors laid the groundwork for standards, business logic, and use case development.

That is essentially the workbench. That are the firms, they have weekly workshops, there are features defined, there are use cases developed, which uses cases are actually pursued and how does that work,

how does the business logic behind that work, which standards are needed. (AU3 – Head of Industry Business Networks, Software Provider)

The project organizations frequently took on a governance role, standardizing processes and certifying components and participants, often through focus committees representing a microcosm of the broader ecosystem. For instance, Auto-Eco initiated a governance body early to ensure longevity beyond public funding.

We started now with [operating company] for the German market, but later, the idea is to have one operating unit and one association for each market. And of course, you need someone to drive that in the new markets, but I don't care if that is a company, like [Automotive OEM], or an industrial association. (AU10 – Head of Digital Production, Automotive OEM)

Both Auto-Eco and Mobility-Eco established infrastructure operating units, with Auto-Eco envisioning multiple units aligned with its decentralized structure and Mobility-Eco distributing ownership across firms. The ownership and control of legal entities like governance bodies and operating units often fell under a conglomerate of actors, setting the stage for the respective ecosystem's governance model.

**Establishing governance mechanisms** was pivotal for steering the ecosystem's progress, delineating boundaries, and instituting rules and sanctions. The degree of openness in governance participation was critical, as it determines who can influence the ecosystem's evolution. This openness shifted over time, e.g., from exclusively funded project beneficiaries to a more open dedicated governance body. For decentralized ecosystems like Auto-Eco, balancing power to avoid monopolies in operation was paramount. Reaching a power balance and reducing the possibility of monopolies, especially in the data ecosystem's operation, were favored.

What a power would this one operating company have? And to distribute this among multiple shoulders is quite sensible. Else we are there, where we currently are with the hyperscalers. (AU2 – Department Head Logistics, Research Institute)

While not all ecosystem members could directly participate in establishing all governance mechanisms, technical components like open-source dataspace connectors and self-sovereign identities ensured some degree of influence on technical standards, aligning with a decentralized or federated governance ideal. The formalization and extent of governance mechanisms were defined and set up in the early phases but continued to evolve. Generally, contractual agreements defined governance participation and sanctions for deviation.

As the large-scale data ecosystems matured from public-funded projects to self-sustaining entities, a strategic shift

towards a more centralized operating model became evident. This shift addressed the need for structured governance and coherent direction, reconciling the initial openness with the practicalities of operation. A centralized governance body emerged with the authority to direct and standardize processes and sanction deviations, while balancing power to prevent monopolistic tendencies. During their inception, ecosystems like Auto-Eco and Mobility-Eco embraced open participation, fostering diversity and power balance.

We try to represent a broad spectrum with the diversification of participants. (AU5 – Product Owner, Automotive OEM)

However, as they matured, the reliance on a more centralized operating company reflected the operational necessity for a structured approach while striving to maintain the decentralized vision through open-source contributions and governance structures. Such an orchestrator enabled sustained operation beyond the volatility of funding cycles, ensuring the ecosystem's evolution through formalized control structures. The decentralized ethos was maintained organizationally by including multiple entities in the governance body and technically through open-source contributions and technical governance measures such as self-sovereign identities.

#### Data ecosystem operationalization of specialized data ecosystems

The specialized data ecosystem cases, Marine-Eco and Aero-Eco, were marked by informal structures and self-regulation, with an apparent inclination towards decentralization and equitable power sharing. However, they also presented more emergent cases of data ecosystem initiatives. Further evolution of the ecosystems will show how they will implement fair governance and evolve in the post-project stage. A transition akin to the mature cases may unfold as they progress to production.

Their **development of an operating model** was influenced by the idea of equitable distribution of power and decentralization within the ecosystem. In the case of Aero-Eco, its decentralized infrastructure had a significant influence on how the operating model was structured.

I would say we are mainly a technology provider, but also, due to the way our technology is structured—specifically with this decentralized aspect—we are shaping how they will think about it in the end. Our technology is more decentralized than other applications. (AE7 – Software Engineer, IT Startup)

However, the decentralized ideals presented challenges when **establishing governance mechanisms**. The specialized data ecosystems oscillated between maintaining

decentralized principles and introducing more centralized management to ensure alignment and operational efficiency.

Whether you build a closed ecosystem within the consortium for the big players, or you build an open ecosystem with the goal of creating an entirely new data economy where everyone can participate, that's the subtle but significant difference in mentality. (AE7 – Software Engineer, IT Startup)

In Marine-Eco, the governance structure included entities responsible for defining governance rules, such as what data can enter the ecosystem.

Someone must be there, and of course, that someone can also be a legal entity, who says, 'Hey, I am the one who represents this data space, and I also define the governance rules for this data space.' So, what comes in, what doesn't come in? (MA1 – SVP Digital Ecosystems, Internet Service Provider)

Ensuring integrity within these ecosystems was critical. In Aero-Eco, for instance, extensive due diligence was carried out to guarantee that participants in the ecosystem were trustworthy. If malicious activity is detected, the ecosystem can trace it back to the company responsible and take corrective actions.

### Dissemination and evolution of large-scale data ecosystem

The conceptualized operational structures were then implemented within the scaling and innovation activities of the ecosystem. This expansion was propelled by the integration of new actors and partners and the execution of go-to-market strategies for the solutions crafted within the ecosystem, all contributing to the formation of a widespread network. Long-term scaling strategies ensured the overall progression of the ecosystem. Scaling the data ecosystems presented a complex endeavor where the decentralized ideal often encounters practical tensions, for example, around issues of mutual trust, equality, and data sovereignty. These principles were fundamental to the ecosystems' integrity and sustainability but could conflict with the pragmatic needs of expansion and integration of new actors.

**Go-to-market activities** became pivotal in expanding the network and championing solutions developed within as data ecosystems transition from emergence to scaling. These activities often leveraged existing B2B relationships established by consortium members to promote the ecosystem within the industry. Industry-specific events and trade fairs served as platforms for introducing the ecosystem to a diverse audience, from large firms to SMEs. Targeted approaches for SMEs addressed potential perceptions of data ecosystems being tailored to the interests of "big

players." It ensured that the value proposition resonated across the board, including addressing SMEs' unique challenges and contributions within sectors like the automotive industry and keeping the entry barriers for SMEs as low as possible, for example, through providing open-source software or simple integration into existing ERP (enterprise resource planning) systems.

That means low barriers to entry, making this software open-source and available for use, free of charge, not requiring major expenditure, which means using existing ERP systems, for example. Everything is important to keep the barriers low for SMEs. (AU5 – Product Owner, Automotive OEM)

Further, the actors that develop software solutions for the ecosystems used their existing distribution structure as go-to-market channels for their data ecosystem solutions. This was mainly relevant for existing customers who already use software solutions from the same provider. For example, one software provider offering ERP software components offered additional solutions for supply chain transparency based on Auto-Eco.

We have our usual distribution structures. So that means we surely also cooperate with our sales crew. Naturally, they approach existing customers. (AU3 – Head of Industry Business Networks, Software Provider)

A deliberate **scaling strategy** was essential for the steady growth and maturation of the ecosystems. Initially, growth was mainly confined within the boundaries of funded projects with a select group of partners. However, opening up to new partners was recognized as necessary to foster a decentralized network's value, which is directly related to its participant count.

That is the way with networks, especially in a decentralized organization, the value is determined by the number of participants. (AU7 – Principal Project Manager, IT Consultancy)

Using partners already established in a market segment allowed scaling operations more efficiently by leveraging their existing reputation and business relationships, including applying some contractual force to push suppliers into the data ecosystem.

It is, of course, also an option to contractually encourage companies to join the ecosystem through supplier agreements and terms and conditions. (AU10 – Head of Digital Production, Automotive OEM)

The development and standardization of infrastructure components within one ecosystem and their contribution to broader standardization initiatives can lower entry barriers and, thus, further aided scaling efforts. Nevertheless,

this development often revealed a centralization of influence, with the evolution of components such as open-source data space connectors being dominated by a few large firms. Despite the open-source nature intended to support a decentralized model, the actual development reflected a concentration of control due to the limited diversity of contributors.

### Dissemination and evolution of specialized data ecosystem

For the specialized data ecosystems, scaling strategies relied more heavily on securing tactical partnerships to enable growth. Unlike large-scale data ecosystems, which could leverage their size and pre-existing relationships to scale, the specialized ecosystems focused on attracting core partners who brought strategic value to the ecosystem.

**Go-to-market activities** for these ecosystems often included traditional dissemination methods, such as trade fairs, technical presentations, and conferences, where the existing ecosystem actors tried to convey their ecosystem's values and benefits.

We use the traditional methods of dissemination, such as presence at trade fairs, demonstrations at the fairs, and technical presentations at conferences. (AE1 – Project Manager, Aerospace Research Institute)

Both specialized data ecosystems emphasized a **scaling strategy** centered around the scalability of their use cases within the ecosystems. As more use cases emerged, the platform attracted further participation, thereby increasing the ecosystem's scale and reach. To achieve this, the ecosystems focused on attracting key partners that bring strategic value to the ecosystem.

We ensure scaling by selecting tactical partners that have strategic contributions to the ecosystem, like [companies A and B] for wind turbines and for energy data. (MA1 – SVP Digital Ecosystems, Internet Service Provider)

The transition from a funded project to an independent, self-sustaining entity was regarded as the key to the long-term success of the specialized data ecosystems. The ability to attract more partners over time through low barriers of entry and scalable infrastructure was pushed to ensure the necessary continued growth to move towards this self-sustaining state.

### Managing the governance trade-off

Overall, these scaling efforts reflected the ecosystems' practical adaptation to the interplay between their decentralized vision and market and industry dynamics realities. As the ecosystems grew, they encountered operationalization challenges that necessitated some degree of centralization to manage complexity and orchestrate the scaling of the ecosystem

initiative. The goal was to nurture an ecosystem that upholds the core values of decentralization while accommodating its participants' diverse and sometimes competing interests. The data ecosystems had to balance the influence of large, established firms with the decentralized ideals that underpin the ecosystem's ethos. The dynamics of open-source development in these ecosystems illustrate this tension. While open-source software is inherently decentralized, the practical contributions were often centralized among a few dominant industry players who were able to afford to let employees spend time on the development. This centralization can dilute the ecosystem's decentralized nature, as the resources and interests of these major contributors steer the majority of development.

The trade-off was explicit for the development and enforcement of standards. The standards supported the decentralized ethos and were developed through working groups and committees. However, their enforcement within and dissemination beyond the ecosystem often required a more centralized authority, as seen in the formation of governance bodies and operating units within entities like Auto-Eco and Mobility-Eco. Thus, while the ecosystems aspired to a decentralized model promoting data sharing and open participation, they had to navigate the centralization-decentralization trade-off to operationalize effectively. The pragmatic shift towards a structured approach to manage governance, orchestrate participant interactions, and maintain system integrity brought about a drift towards centralization, even within fundamentally decentralized frameworks. This evolution reflected the ecosystems' adaptations to the complexity of coordinating a diverse and dynamic network while maintaining the core principles of the decentralized vision.

## Discussion

Our study addresses the emerging tension between the ideal of decentralization in data ecosystems and the operational drift toward centralization due to increasing complexity. Despite the recognized value of decentralization for promoting equity, transparency, and mutual trust (Beverungen et al., 2022; Otto & Jarke, 2019; Scheider et al., 2023), the operational realities often necessitate a trade-off towards more centralized governance structures to manage complexity (Chen et al., 2022; Wareham et al., 2014). This research shows how data ecosystems emerge through and with decentralized ideals. With growing complexity, organizational tensions arise that are mitigated by governance trade-offs and lead to a federated governance approach that balances the decentralization-centralization continuum. The results provide insights into the dynamics of maintaining decentralization principles amid growing complexities, thus contributing to understanding governance structures in evolving data ecosystems.

### Processual structure of the centralization-decentralization trade-off

The inductive analysis of the multiple-case study allows us to conceptualize the evolutionary mechanisms of innovation, adoption, and scaling for digital infrastructures in a data ecosystem context (Henfridsson & Bygstad, 2013). Early research on data spaces developed through the IDS standard, which acts as the technical foundation for the researched data ecosystems, conceptualized them as an alliance-based multi-sided data platform that starts its evolutionary trajectory through the adoption mechanisms and then progresses to innovation and scaling activities (Otto & Jarke, 2019). Our research extends these insights to more mature phases of data ecosystems, which are further along their evolutionary path, and shows the specific dimensions relevant to the evolutionary mechanisms.

The results reveal a processual structure depicted in Fig. 3 that underscores the emergence of data ecosystems propelled by a vision of decentralization. Initially, the formation of these ecosystems is marked by a collective drive towards this decentralized ideal, laying the groundwork for creating technical and organizational standards alongside identifying collaborative opportunities for value creation. Project consortia play a pivotal role in fostering early adoption within their networks and spearheading innovations. Within the early innovation phases, the identification and development of value co-creation opportunities interact with the development of technical and organizational standards by creating requirements that are then addressed and adopted through the standards. Further, the standards are often instantiated through specific use cases.

As these initiatives progress to operationalizing data ecosystems for production and scalability, they encounter operationalization tensions. These tensions stem from the need to balance the decentralized ethos with the practical necessities of centralized orchestration that addresses the growing complexity (Parker et al., 2017; Gawer, 2014). The step towards operationalization inadvertently softens the decentralized ideal, with the emergent operating models creating a drift towards more centralized structures. Despite technical features, such as self-sovereign identity, that inherently resist centralization within the data ecosystem, the practical realization of establishing operating units and other organizational facets creates a discrepancy between the aspiration for a decentralized operating model and the actual implementation. This highlights an underlying tension between the inherent technical characteristics that support decentralization and the organizational strategies required for the data ecosystems to function effectively.

The setup of an operating model and governance mechanisms drives the efforts to disseminate and evolve the data ecosystem, which acts as an innovation and scaling mechanism. Both dimensions also interact. For example, the setup of a robust operating model helps scale the ecosystem through onboarding new partners. These new partners must then be governed, leading to governance mechanisms suitable for the initial alliance members and a wide range of potential ecosystem actors.

In contrasting large-scale data ecosystems, Auto-Eco and Mobility-Eco, with specialized data ecosystems, Marine-Eco and Aero-Eco, a clear differentiation emerges across the processual structure. During the decentralized emergence phase, complex data ecosystems are driven by broad,

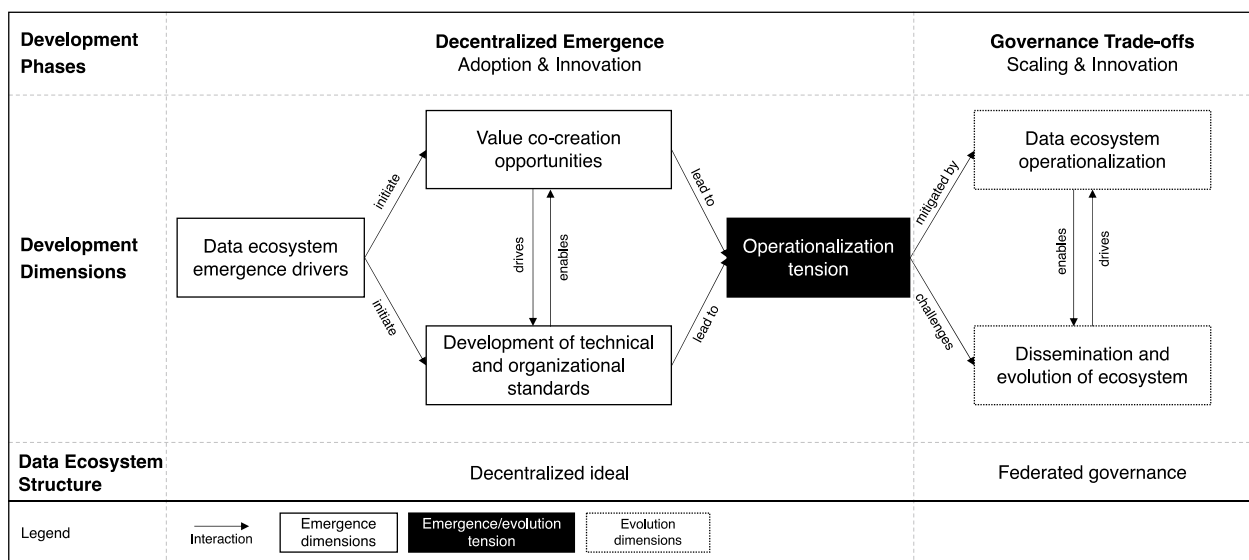


Fig. 3 Processual structure of emergence and development of data ecosystems

ambitious ideation that centers on establishing a democratized, sovereign data-sharing environment. In specialized data ecosystems, the emergence is more restrained, focusing on utilizing existing standards for a select number of use cases. As these ecosystems grow in complexity and orchestration effort, complex data ecosystems often show a growing disparity between their decentralized ideals and the operational reality. This leads to a more significant drift toward organizational centralization to facilitate scaling. Specialized data ecosystems address these challenges through technological means and rely on self-regulation and informal structures, resulting in a smaller drift from their initial decentralized stance and keeping close to the original concept of decentralized data exchange.

As data ecosystems are a nascent practical phenomenon, research on their emergence and evolution is still scarce, especially considering how structure and governance evolve (Kari et al., 2025; Oliveira et al., 2019). We show that the tenets and roles of decentralized governance of data ecosystems change over time. While the decentralized tenet acts as an emergence driver and is crucial for creating a shared vision in the adoption phase through working groups, committees, and similar efforts (Otto & Jarke, 2019), the more mature data ecosystem initiatives do not rely upon full decentralization to innovate and scale the platform as the increasing complexity of the ecosystems requires careful orchestration.

Decentralized governance in data ecosystems faces challenges as they grow more complex (Beverungen et al., 2022). Balancing the ideals of decentralization with the practical need for some central control and orchestration is a delicate act, leading to data ecosystem initiatives managing a centralization-decentralization trade-off. Technological measures like self-sovereign identities and the decision to develop the data space connector as an open-source project reflect a commitment to distributed power. However, creating centralized units within these ecosystems can streamline operations and foster innovation and scaling. This echoes findings in blockchain research, where a mix of central and decentralized governance often works best (Chen et al., 2021). Likewise, the governance of communal resources, such as open-source projects, benefits from a polycentric approach (Ostrom, 2010). This setup allows multiple independent participants to coordinate and govern themselves within a shared framework of rules, leading to ecosystems that serve individual interests and contribute to the collective good (Mindel et al., 2018). Such polycentric governance practices allow the effective steering of the development of an ecosystem where participants also contribute to the development of a wider good and not only their individual benefits, echoing the ideals of the studied data ecosystem initiatives.

## Theoretical and practical implications

Our study advances the theoretical understanding of data ecosystems by addressing the underexplored dynamic between decentralized ideals and the pressures toward centralization that emerge as ecosystems mature. Building on prior research that primarily focuses on the foundational stages of data ecosystems (Gelhaar & Otto, 2020; Oliveira et al., 2019), our contribution to a more nuanced, processual understanding of how governance configurations adapt across different phases of ecosystem evolution is threefold.

First, our findings reinforce and extend the notion of data ecosystems as complex socio-technical arrangements (Möller et al., 2024; Otto & Jarke, 2019), showing that governance tensions are not static but evolve as ecosystems scale and operationalize. In doing so, we contribute to recent efforts to theorize data spaces as meta-organizations (Guggenberger et al., 2025) or network-driven platform ecosystems (Kari et al., 2025). These perspectives conceptualize ecosystems as configurations of multiple, legally autonomous stakeholders that require a negotiated and evolving governance logic. Our study adds to this view by empirically identifying how different ecosystems transition between decentralized and centralized governance, depending on their strategic objectives, participant diversity, and technological maturity.

Second, we extend Kari et al.'s (2025) conceptual distinction between inner and outer governance of the data ecosystem by showing how these layers of governance can shift in emphasis during the evolution of an ecosystem. While early-stage ecosystems may emphasize inclusive, participatory governance mechanisms to foster legitimacy and trust (O'Mahony & Karp, 2022), our study illustrates that maturing ecosystems often require new roles and institutions, such as operating companies or app-store-like intermediaries, that reintroduce hierarchy and control (Beverungen et al., 2022). These shifts demonstrate how ecosystems may need to engage in a trade-off to reconcile conflicting logics: a decentralized structure that promotes fairness and autonomy and a centralized setup that enhances operational viability and scalability.

Third, our findings contribute to a growing literature on platform governance and data sharing (Bodendorf & Bayr, 2025; Fassnacht et al., 2024) by demonstrating that introducing central governance roles is not merely a departure from full decentralization but often a strategic necessity to cope with increasing inter-organizational complexity and manage uncertainty. This complements recent work that highlights how organizational structures and network dynamics co-evolve with data-sharing practices and shows that governance flexibility is not only a managerial challenge but a theoretical lens through which to view ecosystem resilience and adaptability (Fassnacht et al., 2024).

From a practical perspective, our findings offer guidance for organizations building and managing data ecosystems. As ecosystems evolve, governance should not be seen as a static blueprint but as a dynamic capability. Practitioners should deliberately design governance frameworks that allow for flexibility and gradual transitions. This includes building mechanisms that enable role evolution, decision-making adjustments, and introducing orchestration functions without compromising the ecosystem's foundational ideals of decentralization and trust.

Notably, the governance mechanisms should be aligned with the ecosystem's ambition and the scope of its use cases. Ecosystems with broader and more heterogeneous objectives will likely need more formal coordination structures, whereas more focused initiatives may rely on lightweight, goal-oriented governance models. Across both, early efforts to foster legitimacy by increasing participation through inclusive boards or transparent decision-making are crucial to sustain engagement and mitigate future conflicts.

Finally, organizations should treat shared standards and infrastructure as technical enablers and strategic tools for preserving decentralization at scale. Investments in reusable services, common frameworks, and usage policies can support interoperability and trust even as the ecosystem grows more complex and necessitates more central orchestration. In this way, governance becomes a key driver of long-term sustainability and collective value creation.

### Limitations and future research

Our study acknowledges the inherently evolving nature of data ecosystems, which presents certain limitations. Firstly, the dynamism within these ecosystems implies that the derived governance patterns and configurations are subject to continuous change. Therefore, while our findings provide a snapshot, they may only capture part of the trajectory of data ecosystem evolution. Secondly, our data sample focuses on initial members of the data ecosystem initiatives and thus includes a large number of IT and software organizations. A larger sample of manufacturing organizations, especially SMEs, might be needed to create better generalizability of our findings for companies that share and use data in the ecosystems rather than developing components for it.

Future research should focus on a longitudinal analysis to derive patterns and configurations of governance modes and operating models across a broader range of data ecosystem initiatives. While our qualitative approach allows for an in-depth exploration of complex phenomena, quantitative investigations that dissect the development process of large numbers of data ecosystem initiatives will be crucial to verify the results and create a more detailed understanding of their governance structures and operational models. Moreover, future research could draw on the emerging literature on business models in data

ecosystems (Ammann & Hess, 2025) to investigate the impact of governance setups and changes on the business models of ecosystem actors. Combining these dimensions could offer deeper insights into how specific governance decisions not only influence the development and operation of data ecosystems but also the strategies and business models of participating actors (O'Mahony & Karp, 2022).

### Conclusion

Our study sheds light on the transformative journey of data ecosystems as they evolve from decentralized, ideal-driven models to structured governance approaches that balance a trade-off between centralization and decentralization. By conducting a multi-case analysis across different sectors, we identify critical dimensions that influence the governance of emerging data ecosystems. We observe the tension between maintaining the decentralized ideals of equality, trust, and autonomy and the operational necessities imposed by the increasing complexity of these systems. Despite the complexities and challenges encountered as ecosystems mature, the adaptability of governance models indicates a vibrant future where a balance between centralization and decentralization can be achieved. By integrating insights from multiple cases, this research underpins the importance of adaptive governance structures that can withstand the test of time and scale. The positive trajectory of data ecosystems suggests that with continued attention to their evolving governance needs, they can effectively support the democratization of data, drive sustainable growth, and spur new waves of digital innovation.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12525-025-00810-x>.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data availability** The data that support the findings of this study are not openly available due to reasons of anonymity and privacy and are available from the corresponding author upon reasonable request.

### Declarations

**Competing interest** We have no competing interests to disclose. The authors did not receive funding from a research agency.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aaser, M., Kanagasabai, K., Roth, M., & Tavakoli, A. (2020). Four ways to accelerate the creation of data ecosystems. Retrieved 2023-04-13 from <https://www.mckinsey.com/capabilities/quantumblack/our-insights/four-ways-to-accelerate-the-creation-of-data-ecosystems>
- Abbas, A. E., Van Velzen, T., Ofé, H., Van De Kaa, G., Zuiderwijk, A., & De Reuver, M. (2024). Beyond control over data: Conceptualizing data sovereignty from a social contract perspective. *Electronic Markets*, 34, 20. <https://doi.org/10.1007/s12525-024-00695-2>
- Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. *Journal of Management*, 43(1), 39–58. <https://doi.org/10.1177/0149206316678451>
- Ammann, J., Hess, T. (2025). To sell, to donate, or to barter? Value creation and capture through business model types in decentralized data ecosystems. *Electron Markets*, 35, 36. <https://doi.org/10.1007/s12525-025-00775-x>
- Azkan, C., Möller, F., Ebel, M., Iqbal, T., Otto, B., & Poeppelbuss, J. (2022). Hunting the treasure: Modeling data ecosystem value co-creation. In *ICIS 2022 Proceedings* (p. 14).
- Beverungen, D., Hess, T., Köster, A., & Lehrer, C. (2022). From private digital platforms to public data spaces: Implications for the digital transformation. *Electronic Markets*, 32(2), 493–501. <https://doi.org/10.1007/s12525-022-00553-z>
- Bodendorf, F., & Bayr, C. (2025). Shaping platform governance principles to manage interorganizational data exchange. *Information Systems Journal*. <https://doi.org/10.1111/isj.12592>
- Brée, T., Karger, E., & Ahlemann, F. (2024). Shaping the future of data ecosystem research—What is still missing? *IEEE Access*, 12, 103162–103175. <https://doi.org/10.1109/ACCESS.2024.3432969>
- Catena-X Automotive Network. (2025). *Business Areas in the Catena-X ecosystem*. Retrieved 03/04/2025 from <https://catena-x.net/overview-use-case-cluster/>
- Chakrabarti, A., Quix, C., Geisler, S., Pullmann, J., Khromov, A., & Jarke, M. (2018). Goal-oriented modelling of relations and dependencies in data marketplaces. In *Proceedings of the 11th International i\* Workshop @ CAiSE*.
- Chen, L., Tong, T., Tang, S., & Han, N. (2022). Governance and design of digital platforms: A review and future research directions on a meta-organization. *Journal of Management*, 48(1), 147–184. <https://doi.org/10.1177/01492063211045023>
- Chen, Y., Pereira, I., & Patel, P. C. (2021). Decentralized governance of digital platforms. *Journal of Management*, 47(5), 1305–1337. <https://doi.org/10.1177/0149206320916755>
- Constantinides, P., Henfridsson, O., & Parker, G. (2018). Platforms and infrastructures in the digital age. *Information Systems Research*, 29. <https://doi.org/10.1287/isre.2018.0794>
- Curry, E., & Ojo, A. (2020). Enabling knowledge flows in an intelligent systems data ecosystem. In *Real-time linked dataspace: Enabling data ecosystems for intelligent systems* (pp. 15–43). Springer. [https://doi.org/10.1007/978-3-030-29665-0\\_2](https://doi.org/10.1007/978-3-030-29665-0_2)
- De Prieëlle, F., De Reuver, M., & Rezaei, J. (2022). The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of Dutch horticulture industry. *IEEE Transactions on Engineering Management*, 69(4), 940–950. <https://doi.org/10.1109/TEM.2020.2966024>
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550. <https://doi.org/10.2307/258557>
- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32. <https://doi.org/10.5465/amj.2007.24160888>
- Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J., & Satzger, G. (2023). Barriers to data sharing among private sector organizations. In *Proceedings of the 56th Hawaii international conference on system sciences*. <https://doi.org/10.24251/HICSS.2023.453>
- Fassnacht, M., Leimstoll, J., Benz, C., Heinz, D., & Satzger, G. (2024). Data sharing practices: The interplay of data, organizational structures, and network dynamics. *Electronic Markets*, 34(1), 47. <https://doi.org/10.1007/s12525-024-00732-0>
- Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), 1239–1249. <https://doi.org/10.1016/j.respol.2014.03.006>
- Gelhaar, J., & Otto, B. (2020). Challenges in the emergence of data ecosystems. In *PACIS 2020 proceedings* (p. 175).
- Gelhaar, J., Groß, T., & Otto, B. (2021a). A taxonomy for data ecosystems. In *Proceedings of the 54th Hawaii international conference on system sciences*. <https://doi.org/10.24251/HICSS.2021.739>
- Gelhaar, J., Gürpınar, T., Henke, M., & Otto, B. (2021b). Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. In *PACIS 2021 Proceedings* (p. 121).
- Gelhaar, J., Müller, P., Bergmann, N., & Dogan, R. (2023). Motives and incentives for data sharing in industrial data ecosystems: An explorative single case study. In *Proceedings of the 56th Hawaii International conference on system sciences*. <https://doi.org/10.24251/HICSS.2023.454>
- Gieß, A., Möller, F., Schoormann, T., & Otto, B. (2023). Design options for data spaces. In *ECIS 2023 proceedings*.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15–31. <https://doi.org/10.1177/1094428112452151>
- Guggenberger, T. M., Schlueter Langdon, C., & Otto, B. (2025). Data spaces as meta-organisations. *European Journal of Information Systems*, 1–21. <https://doi.org/10.1080/0960085X.2025.2451250>
- HEALTH-X dataLOFT. (2025). *Health-X dataLOFT Platform*. Retrieved 2025–03–30 from <https://www.health-x.org/en/platform>
- Hein, A., Schrieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, 30(1), 87–98. <https://doi.org/10.1007/s12525-019-00377-4>
- Hein, A., Weking, J., Schrieck, M., Wiesche, M., Böhm, M., & Krcmar, H. (2019). Value co-creation practices in business-to-business platform ecosystems. *Electronic Markets*, 29(3), 503–518. <https://doi.org/10.1007/s12525-019-00337-y>
- Henfridsson, O., & Bygstad, B. (2013). The generative mechanisms of digital infrastructure evolution. *MIS Quarterly*, 37(3), 907–931. <https://doi.org/10.25300/MISQ/2013/37.3.11>
- Hsieh, Y. Y., & Vergne, J. P. (2023). The future of the web? The coordination and early-stage growth of decentralized platforms. *Strategic Management Journal*, 44(3), 829–857. <https://doi.org/10.1002/smj.3455>
- Iansiti, M., & Levien, R. (2004). Strategy as Ecology. *Harvard Business Review*, 82(3), 68–78.
- Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8), 2255–2276. <https://doi.org/10.1002/smj.2904>
- Jacobides, M. G., Cennamo, C., & Gawer, A. (2024). Externalities and complementarities in platforms and ecosystems: From structural

- solutions to endogenous failures. *Research Policy*, 53(1), 104906. <https://doi.org/10.1016/j.respol.2023.104906>
- Jussen, I., Möller, F., Schweihoff, J., Gieß, A., Giussani, G., & Otto, B. (2024). Issues in inter-organizational data sharing: Findings from practice and research challenges. *Data & Knowledge Engineering*, 150, 102280. <https://doi.org/10.1016/j.datak.2024.102280>
- Kari, A., Bellin, P., Matzner, M., & Gersch, M. (2025). Governing the emergence of network-driven platform ecosystems. *Electronic Markets*, 35(1). <https://doi.org/10.1007/s12525-024-00745-9>
- Kernstock, P., Altenkamp, P., Böttcher, T., Hein, A., & Krcmar, H. (2025). A configurational approach to understanding data ecosystems. In *Proceedings of the 58th Hawaii international conference on system sciences*. <https://doi.org/10.24251/HICSS.2025.482>
- 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. <https://doi.org/10.2307/249410>
- Lefebvre, H., Flourac, G., Krasikov, P., & Legner, C. (2023). Toward cross-company value generation from data: Design principles for developing and operating data sharing communities. In A. Gerber & R. Baskerville (Eds.), *Design science research for a new society: society 5.0* (Vol. 13873, pp. 33–49). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-32808-4\\_3](https://doi.org/10.1007/978-3-031-32808-4_3)
- Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2021.739>
- Mindel, V., Mathiassen, L., & Rai, A. (2018). The sustainability of polycentric information commons. *MIS Quarterly*, 42(2), 607–631. <https://doi.org/10.25300/MISQ/2018/14015>
- Mobility Data Space. (2025). *Mobility data space – The data sharing community*. Retrieved 2025–03–30 from <https://mobility-dataspace.eu/mobility-data-space>
- Möller, F., Jussen, I., Springer, V., Gieß, A., Schweihoff, J. C., Gelhaar, J., Guggenberger, T., & Otto, B. (2024). Industrial data ecosystems and data spaces. *Electronic Markets*, 34(1), 41. <https://doi.org/10.1007/s12525-024-00724-0>
- Moore, J. F. (1993). Predators and prey: A new ecology of competition. *Harvard Business Review*, 71(3), 75–86.
- Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, 17(1), 2–26. <https://doi.org/10.1016/j.infoandorg.2006.11.001>
- Oliveira, M. I., & Lóscio, B. (2018). What is a data ecosystem? In *dg.o'18: Proceedings of the 19th annual international conference on digital government research: Governance in the data age* (pp. 1–9). <https://doi.org/10.1145/3209281.3209335>
- Oliveira, M. I., Barros Lima, G. F., & Farias Lóscio, B. (2019). Investigations into data ecosystems: A systematic mapping study. *Knowledge and Information Systems*, 61, 589–630. <https://doi.org/10.1007/s10115-018-1323-6>
- O'Mahony, S., & Karp, R. (2022). From proprietary to collective governance: How do platform participation strategies evolve? *Strategic Management Journal*, 43(3), 530–562. <https://doi.org/10.1002/smj.3150>
- Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, 100(3), 641–672. <https://doi.org/10.1257/aer.100.3.641>
- Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the international data spaces case. *Electronic Markets*, 29(4), 561–580. <https://doi.org/10.1007/s12525-019-00362-x>
- Otto, B., Mohr, N., Roggendorf, M., & Guggenberger, T. (2020). *Data sharing in industrial ecosystems*. Retrieved 2023-04-20 from [https://www.mckinsey.de/~media/mckinsey/locations/europe%20and%20middle%20east/deutschland/publikationen/data%20sharing%20in%20industrial%20ecosystems/mckinsey\\_article\\_data\\_sharing\\_in\\_industrial\\_ecosystems.pdf](https://www.mckinsey.de/~media/mckinsey/locations/europe%20and%20middle%20east/deutschland/publikationen/data%20sharing%20in%20industrial%20ecosystems/mckinsey_article_data_sharing_in_industrial_ecosystems.pdf)
- Parker, G., Van Alstyne, M. W., & Jiang, X. (2017). Platform ecosystems: How developers invert the firm. *MIS Quarterly: Management Information Systems*, 41(1), 255–266. <https://www.jstor.org/stable/26629646>.
- Scheider, S., Lauf, F., Möller, F., & Otto, B. (2023). A reference system architecture with data sovereignty for human-centric data ecosystems. *Business & Information Systems Engineering*, 1–19. <https://doi.org/10.1007/s12599-023-00816-9>
- Schleimer, A. M., Jahnke, N., & Otto, B. (2023). Architecture design options for federated data spaces. In *Proceedings of the 56th Hawaii international conference on system sciences*. <https://doi.org/10.24251/HICSS.2023.447>
- Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal*, 50(1), 20–24. <https://doi.org/10.5465/amj.2007.24160882>
- Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc.
- Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Research commentary —Digital infrastructures: The missing IS research agenda. *Information Systems Research*, 21(4), 748–759. <https://doi.org/10.1287/isre.1100.0318>
- Wareham, J., Fox, P. B., & Giner, J. L. C. (2014). Technology ecosystem governance. *Organization Science*, 25(4), 1195–1215. <https://doi.org/10.1287/orsc.2014.0895>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.

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