Engaging in Data Ecosystems: Propositions for Design Characteristics to Foster Value Co-Creation

Marcel Fassnacht
Karlsruhe Institute of Technology
marcel.fassnacht@kit.edu

Carina Benz
Karlsruhe Institute of Technology
carina.benz@kit.edu

Lara Riefle
Karlsruhe Institute of Technology
lara.riefle@kit.edu

Hansjörg Fromm
Karlsruhe Institute of Technology
hansjoerg.fromm@kit.edu

Gerhard Satzger
Karlsruhe Institute of Technology
gerhard.satzger@kit.edu

Abstract

Data ecosystems are gaining traction as key constructs for innovation, collaboration, and co-creation of value. Existing research has predominantly focused on the design of data ecosystems (e.g., incentive mechanisms, governance, or business models) on an institutional level. Surprisingly, actor engagement as the microfoundation for value co-creation has only received very little attention. To advance our understanding of this crucial factor for ecosystem success, we conduct a multiple case study across three real-world data ecosystems. We examine the characteristics of the cases, analyze the observable engagement behavior within each case, and apply a cross-case analysis to derive six propositions for data ecosystem design characteristics that foster value co-creation. Thus, we contribute to unraveling the complex dynamics of data ecosystems with a set of hypotheses that encourage participation and collaboration for more effective value co-creation in emerging data ecosystems.

Keywords: data ecosystem, data sharing, actor engagement, value co-creation, case study

1. Introduction

Data ecosystems have emerged as '[…] 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). In an era of rapid technological advancement and data proliferation, these ecosystems play a pivotal role in harnessing the value of data for innovation, decision-making, collaboration, and value co-creation (Fassnacht et al., 2024; Heinz et al., 2022). The relevance of sharing data in data ecosystems stems from their ability to overcome traditional barriers of data silos and fragmented information landscapes (Fassnacht et al., 2023; Lis & Otto, 2020). By fostering collaboration, data ecosystems offer a holistic approach to leveraging data as strategic assets (Gelhaar et al., 2021a; Heinz et al., 2022). This traction is fueled by the rising legislative attention to creating a regulatory framework for sharing data, such as the Data Act or the Data Governance Act (Jussen et al., 2023).

While the potential for building data ecosystems is immense, one of the remaining challenges is creating value through data sharing and engaging relevant actors – data providers, data consumers, intermediaries, and service providers (Gelhaar & Otto, 2020; Jussen et al., 2024). While value co-creation is a meta-theoretical concept that can hardly be observed or measured (Schüritz et al., 2019), Storbacka et al., 2016) introduced the concept of actor engagement as the micro-foundation of value co-creation. Actor engagement refers to 'both the disposition of actors to engage and the activity of engaging in an interactive process of resource integration within the institutional context provided by a[n] […] ecosystem' (Storbacka et al., 2016, p. 3009). Whether actors engage in data ecosystems may depend on a multitude of characteristics, such as the structure and governance of the ecosystem (Gelhaar et al., 2021a), the ability to derive economic value (Hein et al., 2019; Schreieck et al., 2021), or the technological architecture (Azkan et al., 2020). When actors actively participate, contribute, and collaborate within these ecosystems, they facilitate the flow of data, knowledge, and resources, leading to enhanced value (co-)creation (Storbacka et al., 2016). However, achieving and sustaining high levels of actor engagement can be challenging. This can be observed in emerging ecosystems, such as Catena-X or Cofinity-X, that struggle to evolve from the emergence phase to the expansion phase due to hesitation by their actors to actively engage and share data with other ecosystem participants (Gelhaar et al., 2023; Gelhaar & Otto, 2020). A longitudinal study of the DAMD project (Danish General Practitioners Database) even reports on the collapse of a data ecosystem, partly due to a lack of actor engagement (Aaen et al., 2021). Particularly in the

URI: https://hdl.handle.net/10125/109328
978-0-9981331-8-8

emergence phase of an ecosystem, the overall design and establishment strongly impact the ecosystem's attractiveness and engagement of actors and, thus, ultimately value co-creation (Gelhaar & Otto, 2020; Schreieck et al., 2021). Hence, we pose the following research question:

RQ: What are the design characteristics of data ecosystems that foster actor engagement as a microfoundation for value co-creation?

This paper aims to address this research question by conducting a multiple case study followed by a cross-case analysis and the derivation of propositions regarding beneficial designs of data ecosystems in emergence that foster value co-creation through actor engagement. By applying a structured literature review (Webster & Watson, 2002), we draw a set of characteristics that serve as a construct to characterize three real-world data ecosystems. Through a multiple case study approach (Yin, 2014), we examine three data ecosystems from different industries and with varying scales to identify the actor disposition. Subsequently, we identify the engagement behavior of the actors within these data ecosystems by conducting 15 expert interviews. Through a cross-case analysis of the engagement behavior, we derive propositions for a beneficial design of data ecosystems that foster actor engagement and ultimately value co-creation.

Overall, this research represents a significant step towards unraveling the complex dynamics of data ecosystems and sheds light on the characteristics that foster value co-creation within these ecosystems. By leveraging the insights gained from this study, practitioners can create an enabling environment that encourages active participation, collaboration, and innovation, leading to sustainable and impactful data ecosystem development and value co-creation.

2. Background

2.1. Data ecosystems

The concept of ecosystems is discussed in various contexts to describe interactions among independent organizations in business, strategy, and beyond, such as business ecosystems (Adner, 2017; Moore, 1993), service ecosystems (Vargo & Lusch, 2015), or digital ecosystems (Nachira et al., 2007). Driven by the discourse of big data and the view of data as a strategic asset, the term 'data ecosystem' has recently been adopted in research as a more fine-grained manifestation (Gelhaar et al., 2021a). Thereby, the notion of a data ecosystem translates the broad goals of innovation and value creation from different types of ecosystems into a narrow application context of leveraging the potential of data for innovation and value co-creation in multilateral settings (Oliveira & Lóscio, 2018).

Particularly due to the novelty of the concept, Oliveira and Lóscio (2018) point out the relevance of research on data ecosystems regarding characterization, actor engagement, management, and evolution (Oliveira & Lóscio, 2018). This is further urged in the work of Heinz et al. (2022), who identify three major research streams as the prevalent challenges for research and practice, highlighting in one stream the need for investigating actor engagement in data ecosystems.

Despite the necessity outlined by Oliveira and Lóscio (2018) and Heinz et al. (2022) on fundamental, integral research, existing research on data ecosystems primarily focuses on empirically investigating specific aspects of data ecosystems, such as governance (Lis & Otto, 2021), or data sovereignty (Otto & Jarke, 2019). However, investigating actor engagement in data ecosystems is still nascent in existing literature. Various classification approaches enable the characterization of data ecosystems, such as the morphology of Azkan et al. (2020) or the taxonomy of Gelhaar et al. (2021a), which we can utilize to derive design characteristics relevant to actor engagement in data ecosystems.

However, as highlighted by Gelhaar & Otto (2020), especially the emergence phase of data ecosystems comes with many challenges. In this phase, the data ecosystem aims to transform its objective from emergence to expansion by establishing a business concept that appeals to a large actor base and a scalable infrastructure (Moore, 1993). Particularly, the design of the data ecosystem within this stage heavily determines whether actors are engaging and co-creating value, ultimately affecting its success and potential to thrive (Gelhaar & Otto, 2020).

2.2. Value co-creation and actor engagement

Central to the service-dominant logic (SDL), the concept of value co-creation emphasizes that (economic) value is a result of interaction and resource integration among multiple actors for mutual benefit (Vargo & Lusch, 2004). This value creation takes part within and among service ecosystems, i.e., communities of 'resource-integrating actors [which] are connected by shared institutional arrangements and mutual value creation through service exchange' (Vargo & Lusch, 2015) – a definition that also includes data ecosystems. However, with SDL and value co-creation being meta-theoretical frameworks not directly susceptible to empirical validation, there is a need for less abstract and more context-specific mid-range theories (McColl-Kennedy et al., 2020).

With the concept of actor engagement, this need for a more nuanced perspective on value co-creation is addressed (Storbacka et al., 2016). Informed by the microfoundation movement (Felin et al., 2015), actor engagement follows the idea that collective phenomena (in our case, value co-creation) can only be captured when examining their constituent parts on an actor level (Storbacka et al., 2016). Accordingly, in a three-level model, (1) the abstract idea of value co-creation in service ecosystems on a macro level is broken down via (2) the meso-level view of actors and resources that exercise resource integration patterns facilitated by engagement platforms to (3) the actor engagement on the micro-level. Thereby, the term 'actor' is a generic concept that refers to resource-integrating, service-providing, and value-creating entities (Vargo & Lusch, 2015), including humans, collections of humans (e.g., organizations) or machines (Storbacka et al., 2016). In the context of data ecosystems, actors are mostly referred to as organizations participating in the ecosystem by either sharing data amongst each other (data provider and data consumer) or intermediaries and third parties enabling and facilitating data sharing in ecosystems (Jussen et al., 2023, 2024; Oliveira & Lóscio, 2018). The success of the pursuit of value co-creation by the actors within data ecosystems is thereby determined by the actors' engagement. To capture actor engagement, Engert et al. (2023) propose three engagement behaviors in the context of platform ecosystems. Generating behavior – the engagement of actors by generating and developing the ecosystem, its functionalities, and the actors' activities within the ecosystem. Networking behavior – the interaction and relationships of the actors in the ecosystem and the entire networking within the ecosystem and beyond ecosystem boundaries. Synchronizing behavior – the activities and efforts for aligning and ensuring the fit of the actors in the ecosystem and the ecosystem as such.

Prior work has begun investigating actor engagement in the context of platform ecosystems, e.g., Blaschke et al. (2018) or Schreieck et al. (2021), yet there is no work existing that investigates the impact of the design of data ecosystems in the emergence phase on actor engagement as the observable indicator for value co-creation.

3. Methodological approach

The methodological approach of this work consists of two steps. First, a literature review is carried out to identify the relevant design characteristics to characterize data ecosystems. Second, a multiple case study is conducted that explores three different data ecosystems. We conduct 15 expert interviews to describe these data ecosystems, identify the engagement behavior of the actors within these data ecosystems, and perform a cross-case analysis to derive propositions for a beneficial design of data ecosystems that foster actor engagement and, ultimately value co-creation.

3.1. Literature review

To account for the conceptual knowledge in scientific research, we perform a structured literature review following the approach of Webster & Watson (2002). e aim to identify the characteristics of data ecosystems as the foundation to characterize and compare data ecosystems in a multiple case study. We selected two databases, the AIS eLibrary and Web of Science. To account for related synonyms of the term 'characteristic', we applied the following search string*: ('data ecosystem*') AND (characteristic OR dimension OR taxonomy OR feature OR propert* OR attribute OR criter*).*

Due to the interdisciplinarity of the database Web of Science, we limit our search to the Senior Scholar Basket of Eleven and the JourQual-4 IS outlets and conferences. The literature search reveals an initial sample of eleven papers from the AIS eLibrary and three papers from the Web of Science database. Through an initial screening by title and abstract and a full-text screening, we identify four publications relevant to our research, focusing on existing classification approaches such as taxonomies, typologies, and morphologies. A performed forward and backward search did not reveal additional papers containing novel relevant characteristics of data ecosystems.

To identify the characteristics relevant to our study to characterize the data ecosystems under study, we investigate the classifications regarding similarities and differences of characteristics and derive ten characteristics grouped into four distinct layers. Each characteristic is discussed among the team of authors regarding its relevance and applicability to our study.

3.2. Multiple case study

By investigating three cases of data ecosystems, we follow a multi-case study approach, which matches the comparative nature of our research question (Yin, 2014). The cases are selected to generate 'contrasting results but for anticipatable reasons' (Yin, 2014, p.57). The cases exhibit distinct characteristics while being in the same phase (i.e., the emergence phase) and sharing a common regulatory environment since they all originated within a single country.

Case A is a cross-industry data sharing initiative involving several German multinational corporations. The initiative was initiated to detect data synergies, match data supply and demand, and close data gaps by sharing data (e.g., weather data or forecasting data) between the actors. Case B is a data ecosystem in the process industry initiated and driven by a software company. Within this ecosystem, different manufacturing and process companies can independently share data on machine equipment, machine operations, and maintenance. Case C is a data ecosystem in the automotive industry that aims to increase data-driven interconnectedness along the automotive value chain such as supplier management and inventory alignment or detection of carbon reduction opportunities along the supply chain.

To ensure the validity of our results, we rely on multiple sources of evidence and triangulation (Yin, 2014). We rely on publicly available data (e.g., official brochures, videos of recorded webinars, or public talks) and 15 semi-structured expert interviews with representatives of the three cases. An overview of the cases and corresponding sources is depicted in Table 1.

Table 1. Expert interview sample.

#	Case	Role	Years of experience	Industry	Employees	Interview Duration
I1	A	Manager Business Model Innovation	11-20	Automotive	>100.000	51 min
I2	A	Director IT Innovation	> 20	Consumer Goods	50.001 - 100.000	60 min
I3	A	Manager Digital Consulting	11-20	Chemicals	>100.000	53 min
I4	A	Head Data Assets	11-20	Insurance	25.001 - 50.000	27 min
I5	B	Technical Consultant	11-20	Mechanical Engineering	2.001 - 5.000	30 min
I6	B	Manager Business Development	< 3	Industrial Goods	501 - 2.000	50 min
I7	B	IoT Solution Engineer	6-10	Mechanical Engineering	10.001 - 25.000	45 min
I8	B	Head Digital Unit	> 20	Mechanical Engineering	2.001 - 5.000	31 min
I9	B	Head Digital Services	6-10	Industrial Goods	2.001 - 5.000	34 min
I10	C	Technical Consultant	3-5	IT & Service	25.001 - 50.000	37 min
I11	C	Key Account Manager	11-20	Chemicals	>100.000	47 min
I12	C	Chief Enterprise Architect	> 20	Automotive	>100.000	50 min
I13	C	General Manager	< 3	Automotive	<100	58 min
I14	C	Manager Strategic Partnerships	3-5	IT & Service	100 - 500	56 min
I15	C	Vice President Global Sales & Marketing	11-20	Automation	>100.000	63 min

For the expert interview sample, we follow a purposeful sampling approach based on two criteria (Palinkas et al., 2015). First, experts must have been actively involved in shaping, developing, or working within the respective data ecosystem. Second, technical and business roles are considered to gain insights from various perspectives. We interviewed four experts of Case A (average duration: 48 minutes), five of Case B (average duration: 38 minutes), and six of Case C (average duration: 52 minutes). All interviews are conducted virtually, semi-structured, following predefined guidelines that would help navigate the interview without limiting the natural flow of the conversation and are recorded and transcribed for further analysis. An overview of the expert sample and supplemental information is provided in Table 1.

To characterize the three cases, we analyze the data collected – interview transcripts and supplemental documents – using a deductive coding approach with a nominal category system (Mayring, 2015). The ten characteristics derived from the literature review serve as a foundation and initial code set for the coding process supported by the software MAXQDA. This reveals a total of 837 codes, with the strongest focus on configuration (121 codes), purpose (102 codes), and ecosystem governance (96 codes).

In a second coding cycle, two researchers code the sources following a two-step coding approach. First, open coding is applied to identify the actors' engagement behaviors within the data ecosystems under study. This results in 215 codes. In the second step, we apply provisional coding (Saldaña, 2015) to group the identified codes into pre-defined categories. For this, we rely on the three engagement behaviors (generating, networking, and synchronizing) outlined by Engert et al. (2023) as the applied code book. This results in 114 codes for generating, 51 for networking, and 50 for synchronizing behavior for the three cases, 53 for Case A, 66 for Case B, and 96 for Case C.

To answer our research question, we perform a cross-case analysis (Yin, 2014). The outcome of the coding cycles, the characterization of the data ecosystems, and the engagement behavior of actors serve as the foundation. Based on this, the team of authors discusses the engagement behavior within each case and links it to the corresponding data ecosystem characterization. Discrepancies are mutually resolved through discussion among the team of authors. This allows us to identify the influence of a design characteristic on the engagement behavior in each data ecosystem. Based on this analysis, we identify design characteristics that influence actor engagement positively and derive propositions that account for fostering actor engagement and, thus, value co-creation.

4. Findings

4.1. Characterization of data ecosystems

To build a thorough understanding of data ecosystems, the integral characterization of data ecosystems is an indispensable pillar for the comparison of different data ecosystems and the derivation of propositions on which design characteristics foster actor engagement and thus value co-creation. In the following, we present the analysis of the three cases and characterize these data ecosystems along the ten derived characteristics from the literature review. An overview of characteristics, their descriptions, the sources, and the analysis results of the case characterization is depicted in Table 2. The characteristics are further grouped into

Table 2. Characterization of the data ecosystems.

Layer	Char.	Description	Sources	Case A	Case B	Case C
Structure	Configuration	Actor connections, relationships, and engagement in the data ecosystem	Gelhaar, et al. (2021a); Lis & Otto (2021)	• Alliance-driven formation of actors that are loosely coupled • Collaboration through regular monthly meetings to identify and close data gaps • No aligned strategy and varying motivations • No management commitment	• Keystone-driven formation of actors that are loosely coupled • Collaboration through regular meetings for requirement definition by platform owner • Structure and vision partly present and aligned • Management commitment across all actors	• Alliance-driven formation of actors that are loosely coupled • Collaboration among all actors in various project teams and initiatives • Clearly shared and communicated goal • Management commitment across all actors
	Ecosystem governance	Rules and overarching ecosystem governance regulations ranging from authority-based to trust-based	Azkan et al. (2020); Lis & Otto (2021)	• Decentralized, democratic, trust-based structure • Transparency of governance mechanisms and decisions within the initiative	• Centralized and authority-based structure with a keystone actor (platform owner) • Limited transparency of governance mechanisms and decisions	• Decentralized and democratic structure with the alliance defining general governance • Transparency of governance mechanisms and decisions through association as governing body
Operations	Decision rights	Decision-making on data sharing (centralized by one actor, or decentralized by multiple actors)	Gelhaar et al. (2021a); Lis & Otto (2021)	• No decision rights on ecosystem level • Decentral decisions with democratic structure	• Decision rights defined by keystone actor • Central decision with authoritarian structure	• Decision rights predefined by the alliance • Decentral decisions by actors within the ecosystem guidelines and policies
	Data usage	Usage of data considering data sovereignty and usage policies	Gelhaar et al. (2021b); Lis & Otto (2021)	• No data usage control mechanisms but decentral desired • Usage policies negotiated individually	• Data usage control mechanisms defined and established by keystone actor • Usage policies defined by keystone actor	• Decentral data usage control mechanisms as data sovereignty is a key governance principle • 30 usage policies defined on ecosystem level that are technically enforced in the ecosystem
	Domain	Data origin and environment of data ecosystem emergence	Azkan et al., 2020; Gelhaar et al. (2021b)	• Cross industry (Automotive, Chemicals, Aerospace, Insurance, Finance,…)	• Manufacturing industry	• Automotive industry
	Practices	Transaction relationships between actors encompassing data processes, data sharing design, and additional services	Azkan et al. (2020)	• Individually designed data sharing practices based on use cases • No additional services in place	• Standardized data collection and data access via platform • Services and active collaboration via platform (data analytics)	• Collaboratively enabled use cases with high business relevance through decentralized data sharing • Services (data analytics, authorization etc.)
Architecture	Infrastruct ure	Main technical infrastructure of the data ecosystem	Azkan et al. (202); Gelhaar et al. (2021a); Gelhaar et al., (2021b)	• No specific infrastructure implemented	• Centralized, platform-centric, cloud-based infrastructure • API creation to integrate and connect databases with the platform	• Decentralized open-source infrastructure by utilizing peer-to-peer principles • Usage of data space principles, components, and connectors
	Openness	Accessibility of the data ecosystem; existence of entry barriers or exclusive partnerships	Azkan et al. (2020) Gelhaar et al. (2021a)	• Closed network without expansion endeavors	• Semi-open network with access decision and control by keystone actor • High entry barriers (complex onboarding)	• Open network with open-source infrastructure • Medium entry barriers (limited necessity of IT capabilities and standardized onboarding process)
Economic value	Purpose	Strategic goal of a data ecosystem, the key offering and value generation	Azkan et al. (2020) Gelhaar et al. (2021a); Lis & Otto (2021)	• Exploitation of new business potential, networking, and fostering of collaboration • Enable data-driven use cases and closing data gaps	• Securing market position, new business potentials, increase service level, increase customer experience and relationship • Enable data accessibility and data analytics (increase operational efficiency)	• New business potentials, need for collaboration to close data gaps, increase competitive advantage • Connecting the entire automotive value chain to collaboratively enable use cases with high business relevance
	Value capturing	Value capturing, including payment models and rewards	Azkan et al. (2020) Gelhaar et al. (2021b)	• No value capturing mechanisms defined	• Various potential value capturing mechanisms (monthly subscription fee, pay-per-use,...)	• Multiple value capturing mechanisms (subscription-based, one-time payments, data for data, usage-based,…)

the subsequent four layers: structure, operations, architecture, and economic value. The first layer, *structure*, comprises two characteristics and aims to characterize an ecosystem's overall formation and configuration and its ecosystem governance mechanisms. The second layer, *operations*, clusters characteristics centered around the realization and establishment of data sharing (e.g., decision rights on data access and consumption, practices, and data usage). The third layer, *architecture*, characterizes the technical infrastructure and the degree of openness to participate. The fourth layer, *economic value*, centers around characteristics describing the overall purpose, value creation, and value capturing of the data ecosystems.

4.2. Engagement behaviors in data ecosystems

Next, we assess the actor engagement in the three cases along Engert et al.'s (2023) framework, classifying them into generating, networking, and synchronizing behavior (summarized in Table 3).

In Case A, only limited engagement behaviors are identifiable across all three behavior types. Generating behavior predominantly occurs on an ad-hoc and bilateral basis. Furthermore, there are no established formats for regular exchange and networking at the ecosystem level, resulting in only occasion communication between individual stakeholders from different companies. Synchronization behavior is primarily observed in the efforts of individuals to identify data gaps and cross-industry challenges, with the aim of potentially identifying use cases.

In Case B, the engagement behavior is primarily driven by one keystone actor, the platform owner. Generating behavior is centered around developing and implementing a central platform for sharing data and information, forming the core of the ecosystem. The participating actors, data providers, and consumers focus on digitization as a prerequisite for data sharing and actively share data within identified and developed use cases to address business challenges and generate business value. The keystone actor plays an active role in networking behavior by facilitating matchmaking between data providers and consumers, educating them on platform interactions and usage, and driving the expansion of the ecosystem. Synchronizing behavior is evident through developing and aligning a common strategy and value proposition and standardization efforts such as central contracting and compliance control on the platform.

In Case C, ecosystem-wide generating behavior is observed through the collaborative, agile identification and realization of use cases by cross-company project teams addressing common business problems. This engagement is further depicted by the joint development of a standardized, interoperable, and open-source infrastructure. Intense networking behavior is evidenced by the establishment of regular meetings and active engagement from top management, including participation in the ecosystem's designated board, as

Table 3. Engagement behaviors of actors within the data ecosystems.

Engagement behavior	Case A	Case B	Case C
Generating behavior	• defining use cases bilaterally on technical data models and data gap identification • Processing and sharing data bilaterally and case-by-case • deciding use case development decentrally case-by-case through bilateral negotiations	• identifying and developing use cases based on expected business value in line with revenue models • digitizing assets for usage & sharing on the platform • developing and implementing a central, within-the-ecosystem interoperable platform by the platform owner • active data sharing (provision & consumption) among the ecosystem via central platform	• implementing use cases collaboratively and agilely in cross-company teams with revenue models being individually defined • identifying business problems and defining use cases as a basis for initiating the ecosystem within the automotive industry • implementing a decentral, open-source infrastructure by the actors, including connectors, a data marketplace, a data catalog, and access and verification mechanisms • establishing a technical, fully interoperable infrastructure
Networking behavior	• ad-hoc, individual meetings and workshops among the actors (closed ecosystem) • networking & engagement driven by few individual data enthusiasts with individual interests	• active matchmaking of actors (data provider and consumer) by teams of the platform owner • educating actors on platform usage (by platform owner) • workshops and fair visits for visibility and expansion • pilot projects and prototypes to showcase	• regular meetings by the top management of the key actors • board with regular meetings • engaging and learning through regular exchange with other data sharing initiatives
Synchronizing behavior	• identifying use cases with a data-oriented perspective to resolve different aims & purposes • abstracting industry-specific challenges to identify cross-industry data gaps	• developing and aligning a common strategy among top management of platform owner and actors • jointly elaborating a shared value proposition • establishing a central, standardized contracting and control of compliance	• developing and aligning the ecosystem purpose and strategy among all actors • initiating and funding (financial and resources) by the top management of founding actors • establishing aligned ecosystem governance guidelines, policies, and mechanisms • establishing standardized contracting, including clarification of data sovereignty, access, and usage

well as interactions beyond the ecosystem with other data-sharing initiatives and ecosystems. The high level of engagement is also apparent in synchronizing behavior, with a joint purpose and strategy developed and aligned across the entire ecosystem. Implementation is driven by funding from the top management of founding actors, alongside established ecosystem governance (e.g., policies, guidelines, and mechanisms) and data governance mechanisms (e.g., contracting, data sovereignty, data access, and usage).

4.3. Propositions for design characteristics of data ecosystems that foster value co-creation

By performing a cross-case analysis and triangulating the previous results of our study, the characterization of the cases, and the investigation of the actor's engagement behavior, we derive six propositions that constitute design characteristics of data ecosystems that foster actor engagement as a microfoundation for value co-creation (outlined in Figure 1).

The first proposition reads: '*The establishment of an aligned ecosystem strategy across actors affects value co-creation. It influences top management commitment, funding prospects, and increased alignment of ecosystem purpose and development activities'*. It accounts for the characteristic of configuration and is highly linked with the economic value and purpose of data ecosystems as a crucial component of the strategy. The positive influence of establishing an aligned ecosystem strategy across actors can be distilled from the engagement behaviors within the three cases. I1 (Case A) states that '[company A] pushed the topic because they want to monetize their data and [company B] joined due to personal interest but not primarily for developing disruptive business models.' This highlights the differing aims and purposes further underlined by I4: 'if I don't have a concrete goal […], I'm just looking for new business models and new customer value propositions […], then it becomes very difficult and tends to fail.' In contrast, Case B and C demonstrate the necessity of an aligned strategy as stated by I7 (Case B): 'the management [of the founding actors and the platform owner] has given its full agreement to be active […] and has set common strategic directions and the necessary buy-in'. Also, I10 (Case C) underlines this necessity by quoting, 'there was a very high level of top management commitment which specifies the overall strategy right from the start. Otherwise, this project would not have been possible.'

The second proposition reads: *'The development and establishment of ecosystem governance guidelines, policies, and mechanisms within the ecosystem intensify value co-creation. It creates transparency in decision-making, reduces risks of misbehavior, and allows to incorporate legal requirements and domain-specificities'*. It relates to the characteristic of ecosystem governance. The urgency of a common ecosystem governance becomes evident particularly in domain-specific ecosystems such as Case C with I12 noting 'we have strategic guidance at the portfolio level. And we have established decision-making processes according to certain rules. Otherwise, we would not be able to progress.' Further, I15 (Case C) notes that 'you first have to come together and create the right structures guidelines and policies how to work in the ecosystem. Otherwise, the ecosystem tends to fail.' In contrast, the cross-industry Case A outlines the problems of missing ecosystem governance and difficulties of varying

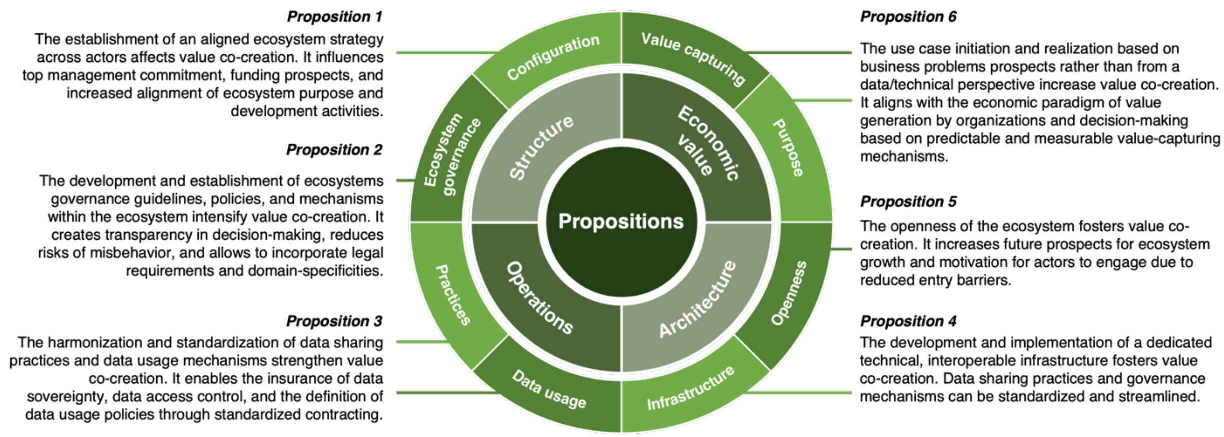

Figure 1. Propositions for design characteristics of data ecosystems that foster value co-creation.

legislation, foci, processes, and governance as I2 states, ‘it was an informal, irregular exchange which resulted in a few use cases that have been bilaterally identified. We did not have any decision-making processes or control mechanisms except individual data sharing contracts.’ While being relevant for domain-specific and cross-industry ecosystems, it is easier for domain-specific ecosystems due to common legislation, guidelines, policies, and mechanisms. It becomes exponentially complicated when combining legislation, policies, and mechanisms from various industries.

The third proposition reads: *‘The harmonization and standardization of data sharing practices and data usage mechanisms strengthen value co-creation. It enables the insurance of data sovereignty, data access control, and the definition of data usage policies through standardized contracting*‘. It accounts for the characteristic of practices and data usage. In Case A, each use case is defined, developed, negotiated, and implemented individually among the participating actors (‘we have bilateral contact with some players on developing a few cases, but this is ongoing, and we still struggle to find consensus on topics like who owns the data, what is done with the data and how can we ensure this by contract.’ (I3)). This imperative of data sovereignty, standardized contracting, and central access mechanisms and control is highlighted by Cases B and C, with I6 (Case B) stating that ‘it is unrealistic that all the contracts are concluded outside the platform, as sharing is now billed on this platform and the time and effort to negotiate these contracts by far exceeds the value of a use case.’ Beyond the contracting, I7 (Case C) emphasizes the need for data sovereignty noting ‘the customer and the market simply demand that they want to have data sovereignty over their data in the network’. Further, I13 (Case C) outlines that ‘data platforms […] have not worked so far for two reasons: firstly, they usually only belong to one central actor, so I don't want to give [data] to someone, and then they're gone. And the second is a lack of data sovereignty. Once I have given the data to someone, I have lost control and don't know what happens to it.’

The fourth proposition reads: *‘The development and implementation of a dedicated technical, interoperable infrastructure fosters value co-creation. Data sharing practices and governance mechanisms can be standardized and streamlined‘*. It addresses the characteristic of infrastructure. While Case B and C implemented dedicated technical interoperable infrastructure, Case A relies on existing infrastructure in the organizations and shares data bilaterally without implementing infrastructure (I1, I3). However, Case B and Case C urge the need for a dedicated infrastructure to foster actor engagement and, ultimately, value co-creation. Case B relies on a central platform initiated and developed by the keystone actor, the platform owner (I8). This is underlined by I7, stating that ‘with a standardized platform, I only have to put in work once for n customers. Otherwise, we would have to coordinate with each customer and find appropriate technical solutions.’ A central platform is further beneficially described regarding actor engagement by I8, noting that ‘organizations trust the company [keystone actor] much more easily to introduce new systems. […] I believe that acceptance would be much lower if we [a participating organization] advertise a system now.’ In contrast, in Case C, a decentralized, interoperable, open-source architecture is developed collaboratively among the actors. As I12 states, ‘this is basically a huge IT project. What do you do if you don't have any infrastructure? Right, nothing.’ In Case C, the open-source characteristic is highlighted as a decisive criterion by I13, outlining that ‘we believe in and want to develop infrastructure open-source. We’re trying to find the best concept based on the masses of people.’

The fifth proposition reads: *'The openness of the ecosystem fosters value co-creation. It increases future prospects for ecosystem growth and motivation for actors to engage due to reduced entry barriers'*. It accounts for the characteristic openness. In Case C, I10 states the positive linkage of openness and open source with actor engagement ('we […] simply didn't want a new platform to be created that belonged to one company, if it is something that we operate cooperatively where many can contribute, it decreases antitrust, entry hesitation, acceptance and overall engagement and contribution of the actors.' Contrary to this, Case B is semi-closed, and Case A is a closed network. In Case A, the closed ecosystem is seen as critical by I1 ('we started with a small number of organizations to reduce complexity, but in my opinion, this also reduces the scope for action and synergies, particularly in a cross-industry setting.'). In Case B, the decision on the participation of new actors is decided authoritarian by the keystone actor, which is seen as critical by participating actors, as underlined by I5 ('we need to reach a critical mass with the platform to make it fly but this is much more difficult if each decision is made by the platform owner.').

The sixth proposition reads: *'The use case initiation and realization based on business problems prospects rather than from a data/technical perspective increase value co-creation. It aligns with the economic paradigm of value generation by organizations and decision-making based on predictable and measurable value-capturing mechanisms'*. It addresses the characteristics of purpose and value capturing and is strongly linked to the first proposition, the operationalization of value capturing and purpose in an aligned strategy. In Case A, use cases are identified with a strong technical focus based on data models and ontologies ('we started to draw data models and look at [each other's] data models if there is something in there that is relevant to us or vice versa.' (I1)). This is particularly criticized by actors I1 and I4 as being not successful due to a lack of the business and legal perspective ('[…] and that was very technical, but that's not nearly enough, because the non-technical the political, strategic and business-related use cases are not included.' (I1)). Further, I4 outlines the necessity of a business problem and the value capturing potential as a decisive factor and starting point to develop use cases ('What challenges do you have [with customers], you need to identify challenges.' (I3); 'you need to show relatively quickly what monetization potential I have and what savings potential I have.' (I4)). Also, actors such as I3 left the ecosystem already 'because it didn't bring any added value for me, so I didn't see any business model for myself.' (I3). This proposition is addressed by Case B and Case C and seen as a critical factor for actor engagement and value co-creation as stated by I8 (Case B): 'you need to get a consistent use case, which then also brings immediate added value and solves a business problem.' In Case C, the ecosystem operates according to the guiding principle: 'I can't solve my business problem, or I see a business opportunity, and that's where [each use case] started.' (I11). In line with this, I15 outlines the need for revenue models ('it's necessary, but it's up to each provider on the platform to decide how to monetize data sharing.' (I15). Further, due to the differences in foci, business understanding, and habits, developing cross-industry use cases is much more challenging than domain-specific cases.

For two characteristics (*decision rights* and *domain*), we found contradictory statements regarding the data ecosystem characteristics and their influence on actor engagement. Consequently, we could not retrieve clear-cut propositions from the gathered data sources. For decision rights, both possibilities, decentral and democratic and authoritarian decision rights, do have advantages and disadvantages (e.g., increased speed and standardization through authoritarian decision rights (I2; I7) vs. fairness and commitment to decisions by all actors through decentralized democratic decision rights (I3; I12). Equally, this is reflected in the investigation of domain-specific vs. cross-industry focus of the data ecosystem. Domain specificity enables a common foundational understanding of the industry and eases the identification of common challenges (I5), while cross-industry ecosystems face difficulties in identifying similarities regarding business problems and data gaps (I2; I14). On the other hand, focusing on a cross-industry setting enables the reduction of hesitation due to competitors not being in the same network and preventing potential antitrust issues (I1; I10).

5. Discussion and conclusion

In this paper, we present six propositions for design characteristics of data ecosystems that foster actor engagement as the microfoundation of value co-creation. For this, we apply a multiple case-study approach (Yin, 2014) investigating three distinct data ecosystems. First, we characterize the data ecosystems under study. As a foundation, we derive characteristics from a structured literature review, which builds the characterization construct. The characterization of the cases along these characteristics is retrieved from fifteen semi-structured interviews. Next, we examine the interviews on the engagement behavior of the actors in the cases following Engert et al.'s (2023) behaviors (generating, networking, and synchronizing). Through cross-case analysis, we triangulate the previous findings to derive six propositions for data ecosystem design characteristics that foster value co-creation.

The propositions are differentiated along the characteristics identified to describe data ecosystems in emergence. We hereby acknowledge that the propositions may be interdependent. While the first (configuration) and sixth propositions (value capturing and purpose) are linked to each other, proposition 2 (ecosystem governance) and proposition 5 (openness) can be contradictory between the degree of openness of an ecosystem and the ability for harmonization and standardization of practices and mechanisms. To date, existing literature focuses on either identifying challenges in the emergence of data ecosystems (Gelhaar & Otto, 2020) or proposing individual design characteristics for business models, ecosystem governance, or incentivization of actors (Gelhaar et al., 2021b; Lis & Otto, 2021; Schweihoff et al., 2022) while neglecting to discuss their findings regarding their impact on value co-creation. In contrast, our approach examines data ecosystems in emergence regarding the design characteristics and the corresponding engagement behaviors (Benz et al., 2024). It triangulates the findings to retrieve concrete propositions for data ecosystem design characteristics fostering value co-creation. With this study, we follow calls for research outlined in various previous studies (Gelhaar & Otto, 2020; Hein et al., 2019; Schreieck et al., 2021). We contribute and urge the need for the conceptualization and design of data sharing ecosystems integrally and enable the triangulation of data ecosystem design with existing literature such as the development of business models (e.g., Schweihoff et al., 2022), incentive mechanisms (e.g., Gelhaar et al., 2023) or ecosystem governance (e.g., Lis & Otto, 2021). Through our strong empirical focus on data sources, we account for the novelty of the concept of data ecosystems, ensuring relevance and applicability in practice while being able to substantially contribute to the emerging research field of data ecosystems. We acknowledge existing work regarding actor engagement and value co-creation in service and platform ecosystems. In addition, we distinguish our work by accounting for the peculiarities of data ecosystems and their various layers of ecosystem structure, operations of data sharing in ecosystems, architecture, and economic value in data ecosystems.

For researchers, our work contributes to the existing body of knowledge in multiple ways. First, we provide valuable propositions for data ecosystems in emergence to foster actor engagement and value co-creation. Each proposition depicts a possibility for future research to investigate the operationalization of this proposition. Additionally, we support researchers in designing data ecosystems with design characteristics to be considered to foster value co-creation within data ecosystems. Our propositions point towards the multidimensionality of data ecosystem characteristics. Consequently, we provide valuable guidelines for developing integral approaches to designing such ecosystems.

For practitioners, our study provides initial guidance and valuable propositions for designing data ecosystems and understanding the influencing characteristics toward fostering value co-creation in data ecosystems. It helps data ecosystem initiatives in the early stage to evaluate their recent endeavors and design against the potential impact on value co-creation and steer towards increasing actor engagement by design. Further, the generalization of the propositions independent of any industry or ecosystem-specific context ensures the broad applicability on various data ecosystem types while providing the ability to specify the proposition in each data ecosystem's context.

As with any study, our research has certain limitations. Case study research relies on the investigated cases and corresponding expert selection. Hence, examining different data ecosystems or modifying the expert sample could result in varying outcomes and propositions. Further, examining the cases and deriving propositions is subject to the chosen characteristics and the selected engagement behavior categories proposed by Engert et al. (2023) which may result in a particular limitation regarding the derived propositions. Additionally, as successfully established data ecosystems are yet sparse, current and future evolutions and advancements of data ecosystems could reveal novel propositions. Further, we aim to introduce generalized propositions and thus may neglect industry-, culture-, or legislation-specific aspects.

While our study lays the foundation for data ecosystem design toward fostering value co-creation, our work presents multiple opportunities for future work. While we investigate three distinct data ecosystems, our work could be extended toward investigating additional ecosystems to evaluate, verify, or refine our propositions. The generalization of our propositions also allows them to be explored and refined in specific contexts, such as within a particular industry or domain. Further, propositions can be interlinked or contradictory, which points toward a future research opportunity to investigate the interrelations and potential affections of the propositions found. Further, the propositions could be transformed into concrete actions in data ecosystem design, and their specific influence on value co-creation in a longitudinal study could be quantitatively and qualitatively observed and assessed. Thus, with this work, we see great potential to contribute to both research and practice. We hope to ease a more targeted design of data ecosystems that fosters value co-creation, ultimately leading to data ecosystems thriving beyond the emergence phase.

6. References

Aaen, J., Nielsen, J., & Carugati, A. (2021). The dark side of data ecosystems: A longitudinal study of the DAMD project. *European Journal of Information Systems*, *31*, 1–25.

Adner, R. (2017). Ecosystem as Structure. *Journal of Management*, *43(1),* 39–58.

Azkan, C., Möller, F., Meisel, L., & Otto, B. (2020). Service Dominant Logic Perspective on Data Ecosystems - A Case Study based Morphology. *ECIS 2020 Proceedings.*

Benz, C., Riefle, R., & Satzger G. (2024). User Engagement and Beyond: A Conceptual Framework for Engagement in Information Systems Research. *Communications of the Association for Information Systems, 54,* 331-359.

Blaschke, M., Haki, K., Aier, S., & Winter, R. (2018). Capabilities for Digital Platform Survival: Insights from a Business-to-Business Digital Platform. *ICIS 2018 Proceedings.*

Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press.

Engert, M., Evers, J., Hein, A., & Krcmar, H. (2023). Sustaining complementor engagement in digital platform ecosystems: Antecedents, behaviours and engagement trajectories. *Information Systems Journal*, *33(5)*, 1151-1185.

Fassnacht, M., Benz, C., Bode, J., Heinz, D., & Satzger G. (2024). Systematizing Data Sharing Practices: A Taxonomy. *ECIS 2024 Proceedings*.

Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J., & Satzger G. (2023). Barriers to Data Sharing among Private Sector Organizations. *HICSS 2023 Proceedings.*

Felin, T., Foss, N., & Ployhart, R. (2015). The Microfoundations Movement in Strategy and Organization Theory. *The Academy of Management Annals*, *9*, 575–632.

Gelhaar, J., Groß, T., & Otto, B. (2021a). A Taxonomy for Data Ecosystems. *HICSS 2021 Proceedings.*

Gelhaar, J., Gürpinar, T., Henke, M., & Otto, B. (2021b). Towards a Taxonomy of Incentive Mechanisms for Data Sharing in Data Ecosystems. *PACIS 2021 Proceedings.*

Gelhaar, J., Müller, P., Bergmann, N., & Dogan, R. (2023). Motives and Incentives for Data Sharing in Industrial Data Ecosystems: An Explorative Single Case Study. *HICSS 2023 Proceedings*.

Gelhaar, J., & Otto, B. (2020). Challenges in the Emergence of Data Ecosystems. *PACIS 2020 Proceedings.*

Hein, A., Weking, J., Schreieck, M., Wiesche, M., Böhm, M., & Krcmar, H. (2019). Value co-creation practices in business-to-business platform ecosystems. *Electronic Markets*, *29(3),* 503–518.

Heinz, D., Benz, C., Fassnacht, M., & Satzger, G. (2022). Past, Present and Future of Data Ecosystems Research: A Systematic Literature Review. *PACIS 2020 Proceedings*.

Jussen, I., Fassnacht, M., Schweihoff, J., & Möller, F. (2024). Reaching for the Stars: Exploring Value Constellations in Inter-Organizational Data Sharing. *ECIS 2024 Proceedings.*

Jussen, I., Schweihoff, J., Dahms, V., Möller, F., & Otto, B. (2023). Data Sharing Fundamentals: Characteristics and Definition. *HICSS 2023 Proceedings*.

Lis, D., & Otto, B. (2020). Data Governance in Data Ecosystems – Insights from Organizations. *AMCIS 2020 Proceedings.*

Lis, D., & Otto, B. (2021). Towards a Taxonomy of Ecosystem Data Governance. *HICSS 2021 Proceedings*

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12th edition)). Beltz.

McColl-Kennedy, J. R., Cheung, L., & Coote, L. V. (2020). Tensions and trade-offs in multi-actor service ecosystems. *Journal of Business Research*, *121*, 655–666.

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, *71*(3), 75–86.

Nachira, F., Dini, P., & Nicolai, A. (2007). *A network of digital business ecosystems for Europe: roots, processes and perspectives*. Publications Office of the EU.

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem? *dg.o 2018 Proceedings*

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

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42(5),* 533–544.

Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd edition). SAGE Publications.

Schreieck, M., Wiesche, M., & Krcmar, H. (2021). Capabilities for value co-creation and value capture in emergent platform ecosystems: A longitudinal case study of SAP's cloud platform. *Journal of Information Technology*, *36(4),* 365–390.

Schüritz, R., Farrell, K., Wixom, B., & Satzger, G. (2019). Value Co-Creation in Data-Driven Services: Towards a Deeper Understanding of the Joint Sphere. *ICIS 2019 Proceedings.*

Schweihoff, J., Jussen, I., Stachon, M., & Möller, F. (2022). Design Options for Data-Driven Business Models in Data-Ecosystems. *Workshop NaWerSys 2022 Proceedings.*

Storbacka, K., Brodie, R. J., Böhmann, T., Maglio, P. P., & Nenonen, S. (2016). Actor engagement as a microfoundation for value co-creation. *Journal of Business Research*, *69(8),* 3008–3017.

Vargo, S., & Lusch, R. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, *68(1),* 1–17.

Vargo, S., & Lusch, R. (2015). Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science*, *44*, 5–23

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26(2),* xiii–xxiii.

Yin, R. K. (2014). *Case Study Research: Design and methods* (5th edition). SAGE.