Government Information Quarterly 43 (2026) 102107

Contents lists available at ScienceDirect

Government Information Quarterly

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

Integrated governance in data ecosystems: A conceptual framework consolidating collaborative and data governance

Federico Bartolomucci [a,*], Edoardo Ramalli [b], Valeria Maria Urbano [a]

[a] Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, Italy
[b] European Commission - Joint Research Centre, Via Enrico Fermi 2749, Ispra, Italy

ARTICLE INFO

Keywords:
Data governance
Data ecosystem
Integrated governance
Collaborative governance
Qualitative research

ABSTRACT

The potential benefits deriving from inter-organizational data sharing have increased over time, leading to an intensified interest in data ecosystems. The governance of these endeavors depends on both collaborative and data governance dimensions. However, previous research has often treated these dimensions separately, creating silos that hinder the capacity to deliver value considering their socio-technical nature. Addressing this gap, this study investigates the intertwined relationship between these two dimensions within data ecosystems. It does so by questioning which existing and most relevant relationships exist between them, as well as the nature of these relationships. To this end, we adopt a multiple case study approach, analyzing five data ecosystems. The research led to the development of a conceptual framework for Integrated Governance, highlighting the need for a holistic socio-technical approach that addresses collaborative and data governance dimensions as intertwined. The framework unveils 24 core relationships between these dimensions in data ecosystems and provides insights on the nature of the relationships, distinguishing among causal, explanatory, concurrent, chronological, and overlapping ones. This work introduces a new perspective in the academic discourse on data sharing providing actionable insights for practitioners and enabling them to design and manage data ecosystems more effectively.

1. Introduction

The rapid growth in data volume and the increasing need to extract value from it have intensified scholarly interest in Data Ecosystems (DEs). DEs are often used as an umbrella term to describe multi-actor arrangements for inter-organizational data exchange. Oliveira and Lóscio (2018) defined a DE as a loose set of interacting actors that directly or indirectly consume, produce, or provide data and other related resources (e.g., software, services, and infrastructure) in which actors are connected through diverse relationships and simultaneously collaborate and compete. Different types of collaborative environments have been conceptualized as DEs (Oliveira, Barros Lima, & d. F., and Farias Lóscio, B., 2019), including data collaboratives (Susha, Grönlund, & Van Tulder, 2019; Verhulst & Sangokoya, 2015), data cooperatives (Bühler et al., 2023). Despite terminological and operational differences, these collaborative environments share the common aim of generating value through inter-organizational data sharing.

Literature has so far recognized the complexity of these collaborative environments, which lies in their socio-technical nature (Ruijer, 2021), shaped by the interconnection and co-existence of organizational and technological factors (Fang, Zhao, & Li, 2024; Gelhaar, Groß, & Otto, 2021; Lnenicka et al., 2024; Susha, Janssen, & Verhulst, 2017). Accordingly, data-related aspects should be considered jointly with the organizational systems through which data are collected, managed, and used, while acknowledging the central role of people and organizations in these processes (Janssen, Brous, Estevez, Barbosa, & Janowski, 2020). Important contributions in this direction have been provided by Janssen et al. (2020) and Ruijer (2021). Janssen et al. (2020) proposed a comprehensive definition of data governance that explicitly accounts for the actors responsible for the creation and use of data and algorithms and adopts a systems-based, multi-organizational perspective. Ruijer (2021) instead, focuses on data collaborative governance, identifying organizational, political–policy, and technical dimensions as key components shaping governance.

Despite the emphasis placed on both dimensions, the relationship and influence of organizational and technical factors remains insufficiently understood (Möller et al., 2025), with limited efforts devoted to their integration. The literature review by Schmeling, Al Dakruni, and Mergel (2025) confirmed that fragmentation persists in the field, showing that technical architecture decisions are strongly shaped by

* Corresponding author.
E-mail addresses: federico.bartolomucci@polimi.it (F. Bartolomucci), edoardo.ramalli@ec.europa.eu (E. Ramalli), valeriamaria.urbano@polimi.it (V.M. Urbano).

https://doi.org/10.1016/j.giq.2026.102107
Received 22 July 2025; Received in revised form 13 January 2026; Accepted 19 January 2026
Available online 23 January 2026
0740-624X/

organizational structures and calling for future research to more explicitly address these interdependencies.

Being interdependent, these factors alternatively reinforce or threaten each other (Micheli et al., 2023). Their inter-play shapes the ability to derive value from data (Yeh, Eden, Fielt, & Syed, 2025), design viable business models (Susha, Flipsen, Agahari, & Reuve, 2020), enable data interoperability (The New Hanse Project, 2023), manage trust concerns and overcome data-access hurdles (Otto & Hompel, 2022). Treating them in isolation spurs misalignment and can undermine a partnership's capacity to deliver value to its stakeholders (Bartolomucci & Bartalucci, 2024), whereas alignment and coherence among them strengthen the robustness of the DE setting (Micheli et al., 2023).

To substantiate the need for adopting an integrated approach addressing the two dimensions as intertwined, this work provides a general framework on DE governance that integrates the two dimensions defined as Collaborative Governance (CG) and Data Governance (DG). In the context of data sharing, CG can be defined as “the processes and structures of decision-making and management that engage people constructively in data-driven activities across the boundaries of public agencies, levels of government, and/or the public, private, and civic spheres” (Ruijer, 2021), while DG can be seen as “the exercise of authority and control over the management of data” (Abraham, Schneider, & Vom Brocke, 2019). Adopting this holistic perspective, this study addresses the following research questions:

- *RQ1: What are the most relevant relationships between Collaborative and Data Governance factors within DEs?*
- *RQ2: What is the nature of the core relationships between Collaborative and Data Governance aspects within DEs?*

This work addresses these research questions by exploring, through a multiple case studies approach, the existing relations among CG and DG factors in five DEs operating in multiple sectors (e.g., health, logistics, research). We do so under a socio-technical perspective and through a multiple case study. This approach allowed us to develop an interpretative qualitative analysis to deepen the understanding of the identified relationships (Adu, 2019).

Adopting a holistic perspective on DE governance that integrates technological and organizational dimensions, our research puts forward a conceptual framework, displaying the concept of *Integrated Governance (IG)*, intended as *the integration, in a unified perspective, of collaborative and data governance aspects and the consideration of their chronological, causal, explanatory, overlapping and concurrent relations as key determinants of the governance model's effectiveness*. The framework unveils the core relationships between collaborative and DG dimensions in DEs. To complement the analysis, and following the framework developed by (Dey, 2003), we conducted an interpretative analysis aimed at isolating the intrinsic meaning of each of the core relationships identified. The analysis brought us to distinguish five types of relationships: concurrent, chronological, overlapping, explanatory, and causal, each having a different implication for theory and practice. This work contributes to the on-going academic discussion on DE governance (Geisler et al., 2021; Lis & Otto, 2020; Ramalli & Pernici, 2023) and the broader discussion on the governance of data-centered collaborative endeavors such as data collaboratives (Bartolomucci & Leoni, 2024; Ruijer, 2021) and public data ecosystems (Lnenicka et al., 2024; Nikiforova, Lnenicka, Luterek, Milic, & Rodríguez, 2025). The integrated perspective gained through the research can support practitioners in better designing and managing their collaborative environments, identifying the effects of the positive interplay between the dimensions, and avoiding misalignment.

The structure of the paper is as follows. Section 2 discusses the relevant literature and presents examples of recent frameworks that include both organizational and technological dimensions. Building on this foundation, Section 2 introduces the adopted research framework, highlighting its main differences from previously proposed frameworks and outlining the two macro-dimensions of analysis in separate subsections (i.e., CG and DG). Section 3 introduces our methodology, describing the data collection and data analysis steps and explaining the analytical steps applied. Section 4 shows the results and provides an explanation of the core relations identified. In Section 5, we discuss our findings and contributions to the literature, as well as limitations and possible future research directions. Section 6 draws the conclusions of the study.

2. Conceptual background

DEs are data sharing platforms defined as “distributed, open, and adaptive information systems with the characteristics of being self-organizing, scalable, and sustainable” (Oliveira & Lóscio, 2018). A DE shares data from a *data producer* to a *data consumer*, typically through web-services, managing the hosted data and increasing their value. (Oliveira & Lóscio, 2018). The concept of DE, includes a multitude of data sharing environments that can be distinguished based on the policy adopted to manage the data, the DE goal definition, the degree of participant interaction, and the degree of data exchange within the ecosystem (Curry & Sheth, 2018). They can be alternately open or closed in terms of the ecosystem's accessibility; however, data sharing within DE is most of the time driven by economic interests (Oliveira & Lóscio, 2018). Within this landscape different forms of DE have taken shape encompassing data cooperatives (Bühler et al., 2023), city data ecosystems (Liva et al., 2023). More recently, public data ecosystems emerged as socio-technical complex networks where actors interact to find, archive, publish, consume, or reuse data sourced from or funded by public bodies (Nikiforova et al., 2025). Their overarching goals include fostering innovation, enabling sustainable and citizen-centric service delivery, supporting new business development, and creating public value (Degen, Lutzens, Beschorner, & Lucke, 2025; Fang et al., 2024).

Several studies have listed the design principles of DE's governance models (Otto et al., 2017; Otto & Jarke, 2019) and architectural components (Demchenko, De Laat, & Membrey, 2014), particularly in industrial settings for industries' digitalization (Otto, Hompel, & Wrobel, 2019). Governance models may differ based on organizational factors (e.g., Communication modes, institutional logics), political and policy-related elements (e.g., processes transparency), and technical aspects (e.g., type of data aggregated, data sharing infrastructures) (Ruijer, 2021). However, in practice, each DE has its challenges requiring specific customization (Cappiello, Gal, Jarke, & Rehof, 2020). If not properly addressed, these challenges can completely preclude the adoption of such a promising model, and as a result, such data-sharing initiatives fail quickly due to the low adoption or interest, such as in the case of the most recent personal data markets (Scheider, Lauf, Geller, Möller, & Otto, 2023).

Despite the industry-specific and case-dependent nature of these governance-related challenges, several recent frameworks examining DE and, more broadly, inter-organizational data-sharing initiatives, have incorporated both organizational and technological dimensions.

Susha et al. (2017) provided a taxonomy for data collaboratives comprising two layers: a collaboration layer and a data layer. In their empirically and theoretically grounded framework, they distinguished between the supply and demand sides of data collaboratives, identifying different forms of collaboration based on multiple dimensions situated both on the organizational side and on the data side. Their study highlighted the complex interdependencies that characterize data collaboratives, demonstrating the need for analyses that simultaneously incorporate collaboration and data-related aspects.

Gelhaar et al. (2021) argued that the data collaborative framework lacked important characteristics of the ecosystem perspective, such as co-evolution and organizational structures. They proposed a taxonomy consisting of three meta- dimensions: Economic, Technical, and Governance. While the Technical dimension addresses the characteristics of the technical architecture of the DE, the Economic dimension includes dimensions related to the business model perspective (including the domain, purpose, and organization of the DE), and the Governance dimension covers aspects of data ownership and actor dependencies within the DE.

Building on Yang and Maxwell (2011); Pardo, Gil-Garcia, and Luna-Reyes (2010), Ruijer (2021) examined living labs and proposed a holistic framework for data collaborative governance, identifying organizational, political and policy, as well as data and technical dimensions as key components shaping governance. In their study, the authors emphasized the interrelations among these elements, offering an overview of the decision-making processes and structures that should be considered when designing and implementing a data collaborative.

More recently, Lnenicka et al. (2024), in their multi-country empirical study on open DEs adopted a three- dimensional framework building on (Gil-Garcia & Sayogo, 2016) including a managerial and organizational context (e.g., governance, financial resources), a political and institutional context (e.g., enabling legislation), and an information and technological context (e.g., technical infrastructure and security standards).

Our framework extended the definition of the technical dimension proposed by Gelhaar et al. (2021) and Gil-Garcia and Sayogo (2016) to incorporate a broader set of technical aspects; we refer to this first dimension as DG. We also refine the Governance dimension, as presented by (Gelhaar et al., 2021), by integrating organizational components, and we refer to this second dimension as CG. At the same time, by observing the cases within a specific time frame, we treat the domain and purpose of the DE as fixed. This does not imply neglecting the dynamic and evolutionary nature of DE (Bartolomucci & Leoni, 2024); rather, it allows us to concentrate on governance intricacies within a window of stable purpose. The same reasoning applies to the institutional context and political elements: although its variations may exert external pressures on the DE, we consider it a broader, macro-level element within which the DE operates at a given point in time. Furthermore, since all the cases analyzed are situated within the same national setting (Italy) and thus do not display substantial institutional variation, we excluded this dimension from our framework. Finally, having considered the original conceptualization of DE from (Oliveira & Lóscio, 2018), we refer our analysis to closed DE, i.e., environments to which only authorized actors can access, and thus do not include the degree of openness among the dimensions of the framework.

In the following paragraph we describe the research background that brought us to isolate the key component of CG and DG to be included in the framework. In doing so, as noted by (Schmeling et al., 2025), we run into the fragmentation of research streams characterizing the field, which creates inconsistencies in the terms used to refer to the same or similar concepts. Thus, to isolate DG and CG aspects composing the framework, we had to go through multiple iterations on the meaning of each term, till agreeing on a specific ontology for each term. This process aimed at avoiding as much as possible semantic overlaps among different dimensions, causing confusion and correlations between the variables. For instance, data sharing and data exchange have been used as synonyms in both CG and information systems research domains, although their meanings have evolved throughout time (Jussen, Schweihoff, Dahms, Möller, & Otto, 2023). The iterative process of converging on the most salient collaborative and DG dimensions facilitated the development of a shared vocabulary. By developing a common vocabulary, we created a unified interpretive frame that and prevented divergent understandings of the dimensions under analysis among interviewees. The following paragraphs discuss the background used to define the framework components, while Appendix A provides definitions of each dimension.

2.1. Collaborative governance

Literature on CG has extensively discussed the importance of governance frameworks for the success of collaborative environments (Agranoff, 2006; Ansell & Gash, 2008; Emerson, Nabatchi, & Balogh, 2012). Bryson, Crosby, and Stone (2015) conceptualized governance as a complex construct that goes beyond the interaction of processes and structures and considers, for the first time, elements such as technology and outcome measurement as key components of the collaborative regime. CG models often distinguish different components of the collaborative regime. Among them, literature agrees on identifying elements describing the antecedents or initial conditions (i.e., the socio-cultural and political conditions in which the partnership takes place); structures (i.e., formal elements that constitute the backbone of the collaborative effort like agreements and governing bodies); processes (i.e., elements of interaction among the partners, including communication, trust and informal leadership) (Bryson, Crosby, & Stone, 2006); and outcomes (i.e., impacts on partners and the external environment).

Building on previous literature on collaborative governance (Agranoff, 2006; Ansell & Gash, 2008; Bryson et al., 2006; Bryson et al., 2015; Emerson et al., 2012), recent studies and adapted CG models in the realm of data sharing environments (Micheli et al., 2023; Oliveira et al., 2019; Ruijer, 2021; Susha, van den Broek, van Veenstra, & Linåker, 2022). According to these perspectives, data-sharing ecosystems' governance models are constructs in which social, technological, and contextual components interact and determine the success of the partnership. CG assumes, therefore, the characteristic of a context-dependent, dynamic and socio-technical construct (Bartolomucci & Leoni, 2024) that transcends the dichotomy of technical aspects of DG interacting with organizational aspects of CG but addresses these elements as entangled in a continuously evolving relation.

To isolate organizational governance aspects that could be considered relevant in DE, we have reviewed the most recent literature on organizational aspects of DE (Liva et al., 2023; Micheli et al., 2023), data collaboratives' governance (Bartolomucci & Bartalucci, 2024; Ruijer, 2021; Susha et al., 2022), and their antecedents on CG (Ansell & Gash, 2008; Bryson et al., 2006; Emerson et al., 2012). The review resulted in the selection of eight CG dimensions included in the framework: Trust (TR), Incentives (IN), Competence (CO), Flexibility and Adaptability (FA), Organizational Structure (OS), Decision Making (DM), Funding Model (FM), Community & Stakeholders' engagement (CE). Their meanings are described in Appendix A, these are referred to hereafter as *collaborative dimensions* or aspects. Other dimensions, such as leadership, institutional context or formal agreements, that are traditionally part of the CG concept were excluded from the research framework due to either their levels of overlap with the dimensions considered or a lack of preliminary evidence in the literature about their relationship to DG features.

2.2. Data governance

Data governance is identified as a comprehensive framework for assigning decision-making rights and responsibilities across the organization to manage data as a valuable corporate asset effectively (Donaldson & Walker, 2004; Otto, 2011). Despite this broad definition, the concept has been addressed in the literature with varying interpretations and differing boundaries. Alhassan, Sammon, and Daly (2016) have described the relationship between DG and data management, emphasizing the strong connection between the two. *Governance* defines the decisions that need to be taken to ensure effective management and use of resources. *Management* regards the implementation of such decisions.

Over the years, different frameworks and methodologies have been proposed to identify DG activities, *dimensions* or *aspects*. The HORUS identifies five dimensions that concern how to *Hold* information securely and confidentially, *Obtain* information fairly and efficiently, *Record* information accurately and reliably, *Use* information effectively and ethically, and *Share* information lawfully and appropriately (Donaldson & Walker, 2004). Panian (2010) instead, proposed a framework for DG that examines the typical data challenges in a business-oriented manner. The framework includes aspects related to the accessibility, availability, quality, consistency, security, and auditability of the data. Khatri and Brown (2010) identified five *decision domains* that should be considered in the definition of a DG framework: data principles, data quality, metadata, data access, and data life cycle. Data principles designate the guidelines for the use of data assets, setting the organization's standards for data quality. These standards shape how data is interpreted (through metadata) and accessed (data access) by users. Lastly, decisions about the data lifecycle, which includes the production, retention, and retirement of data assets, are crucial to translating data principles into actionable practices within the IT infrastructure (Khatri & Brown, 2010).

More recently, Abraham et al. (2019) argued that, despite the growing importance of DG, the current view on this topic is still fragmented. Lis, Gelhaar, and Otto (2023) discussed instead the evolution of the concept of DG pointing out the necessity of developing more research in relation to the increasing inter-organizational nature of DG. Similarly Möller et al. (2025) affirmed the need to re-examine the foundations of DE, among which DG factors, in light of the recent evolving use cases like digital data and emerging regulations. Nikiforova et al. (2025) positioned data governance models as key research areas to foster the development of public data ecosystems. In this evolutionary context, to identify the relevant aspects of DG for this work, we review the literature starting from the definition and conceptualization of DG (Ladley, 2019), its policies (Salnitri, Ramalli and Pernici, 2024), key elements (Abraham et al., 2019), and principles (Khatri & Brown, 2010) and complemented it with more recent developments in the context of DE (Lis & Otto, 2020; Möller et al., 2025). The review brought to the selection of nine dimensions: Data Collection & Transformation (DC), Data Quality (DQ), Data Accessibility & Findability (DF), Data Security (DS), Data Privacy and Confidentiality (DP), Data Modeling (DM), Data Transparency (DT), Data storage & infrastructure (DI), Data Auditing and Monitoring (DA).

Overall, the isolation of the dimensions of analysis of CG and DG (Dey, 2003) contributed to the definition of the research framework, displayed in Fig. 1. The framework is complemented by the six types of relations, identified by previous research (i.e., concurrent, chronological, overlapping, explanatory, causal and embedded) (Dey, 2003). As further explained in the methodological section, these guided us in interpreting the conceptual nature of the relationship identifies.

3. Methodology

This research adopted an interpretative multiple-case study design, based on an inductive approach (Yin, 2018). The choice of an inductive multiple case study design aligns with our objective of examining contemporary phenomena and processes in real-life contexts (Yin, 2018), enabling us to extrapolate knowledge from empirical evidence while providing an in-depth analysis of the relationships between CG and DG factors. Analyzing DEs in different sectors allowed the study to cover a range of contexts and domains, contributing to the understanding of a specific phenomenon across sectors (Eisenhardt, 1989). Equally, it enabled the triangulation of different perspectives and capturing recurring patterns, thereby ensuring that our conclusions are not limited to the idiosyncrasies of a single context but instead reflect broader dynamics of CG and DG practices. The interpretative approach enabled an exploration of the meanings and perspectives that participants assign to the phenomenon, allowing us to gain insights into the subjective experiences and interpretations within each case (Walsham, 1995).

The selection of cases for this study was guided by two primary criteria. First, we selected DEs that are operational, rather than in the design phase. Second, case selection was contingent on the availability and willingness of key stakeholders to participate in interviews and share their insights. In total, we have analyzed five case studies and developed thirteen interviews, with at least two and a maximum of three interviewees for each case. Interviewees' roles and cases are reported in Table 2. The number of cases was established through a qualitative saturation principle (Baker, Waterfield, & Bartlam, 2018), indicated by the lack of emergence of new significant relationships with respect to those already identified. The analyzed cases are operating in different sectors (health, logistics, research, policies and innovation) and are alternatively promoted by private and public actors. Table 1 provides more details for each case.

The overall research process is depicted in Fig. 2 and it is described in detail below. The process is divided into two main phases: data collection (Section 3.1) and data analysis (Section 3.2).

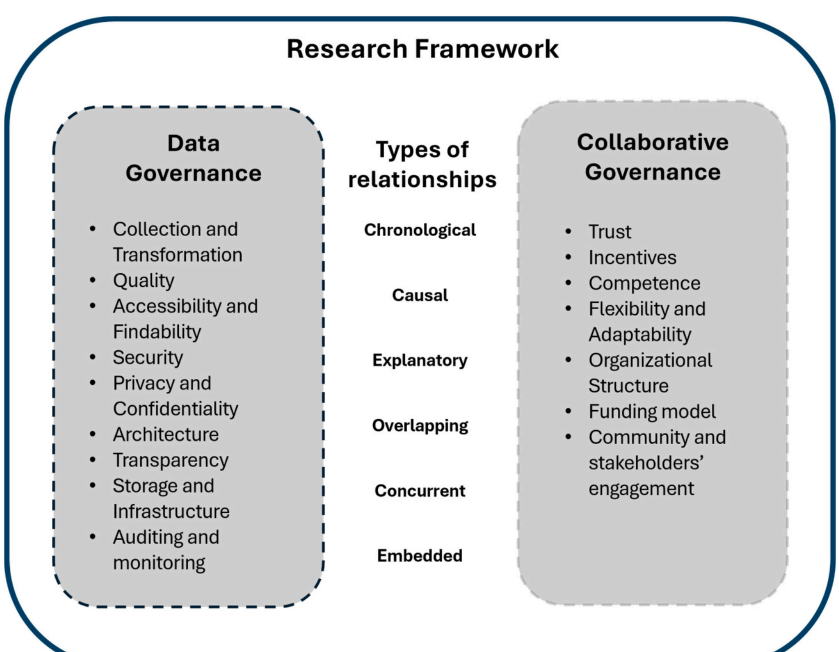

Fig. 1. Research framework.

Table 1
Data ecosystems, sectors, and descriptions.

Data Ecosystem	Sector	Description
SciExpeM	Research	SciExpeM is a research initiative that aims to collect all chemical kinetic data in a unique data ecosystem. These data include ex- pensive scientific experiments as well as predictive models. Sharing data between the research community helps to optimize resources and boost research to reach the carbon-free goal.
Impact Deal	Innovation & Entrepreneurship	Impact Deal is a program for European enterprises aimed at generating impact on society and the environment. It boosts growth through access to datasets provided by public and private partners and leveraging a network of investors, mentors, and stakeholders.
E015	Public Policies	The E015 Digital Ecosystem is an initiative promoted by Regione Lombardia in collaboration with Confindustria, CCIAA of Milan, Confcommercio, Assolombarda, and the Union of Commerce, with the technical-scientific coordination of Cefriel. The E015 Digital Ecosystem fosters the creation of digital relationships between various public and private stakeholders who are interested in enhancing their digital assets by sharing them or enriching their software solutions for their users with functionalities and information shared by other participants.
Malpensa Smart City	Logistics	The Malpensa Smart City of Goods digital ecosystem is a network of ICT-based services that allows all stakeholders in the import-export supply chain to use, integrate, and share relevant information in the handling of air freight arriving and departing from Malpensa airport. The ecosystem aims to facilitate information exchange between the various supply chain actors, providing services that have minimal impact on the information systems used for their commercial and administrative purposes.
Health Big Data	Health	Health Big Data is the Ministry of Health's multi-year MEF-funded project to create a technology platform for collecting, analyzing and sharing clinical and scientific patient data. It aims at developing an integrated and federated Cloud Platform for the collection, sharing, and advanced analysis (AI) of real-time clinical-scientific data of hundreds of thousands of patients from Italian research hospitals (IRCCS). The objective is to create a universal knowledge resource in medicine, based on respect for patient privacy and free access for doctors, scientists, and patients.

3.1. Data collection

The primary data source for our study consisted of semi-structured in-depth interviews conducted with key stakeholders in each case (Table 2). Secondary sources included publicly accessible information available on the websites of the selected cases and sources provided by people interviewed. They were also analyzed to gain contextual understanding of the data collected. The primary data collection was developed through an offline interview followed by a live interview.

Offline Interview. Before the interview, a document containing definitions of DG and CG dimensions was sent to the interviewee to align the terminology (see Appendix A). Interviewees were asked to fill the *Framework Table*, whose rows report the CG dimensions and the columns report the DG dimensions (See Fig. A5 in Appendix A). The *Framework Table* accounts for 72 possible relationships. Participants indicated the presence or absence of a relationship between two dimensions by assigning 1 or 0, respectively. Participants were also asked to highlight the most critical relationships by assigning a value of 2. The order of CG and DG aspects in *Framework Table* was shuffled in each interview. Interviewees filled the *Framework Table* two times: first, focusing on the influence of CG dimensions on DG dimensions, and second, focusing on the influence of DG dimensions on CG dimensions. This activity resulted in two *Draft Interviewee Table* for each interviewee. This bidirectional design enables the identification of asymmetries and reciprocal dynamics between the two governance domains.

Interview. The preliminary offline interview allowed focusing the *Interviewee* on arguing the rationale behind these relationships and sharing relevant experiences. Interviews were structured according to the relationships identified as part of the offline interview. Based on this, we have requested each interviewee to comment on all identified relationships, beginning with the most significant ones and then providing a rationale for all of them. The discussion.

was facilitated through explicative and exploratory questions like: “why did you choose this relationship? Can you explain to us the rationale behind your choice? What is your experience with this relationship?” To reduce bias in the interview directionality, each interview was conducted by at least two authors. During each interview, as a result of the discussion, relations were added, removed or changed in score of importance, ultimately resulting in two *Reviewed Interviewee Table*. Interviews were recorded and transcribed; anonymity and confidentiality were guaranteed to the interviewees to increase the integrity of the responses.

3.2. Data analysis

Data analysis involved two key phases: summarization and filtering, where the data was organized and refined, and coding and interpretation, where the transcripts were analyzed to derive meaningful insights. The analysis aimed to identify robust and recurrent dimensions across heterogeneous use cases, rather than to draw case-specific conclusions.

Summarization & Filtering. At this stage, the input *reviewed interviewee tables* was aggregated by harmonizing and sorting their axes, then an element-wise sum was performed to generate a composite *importance score* for each relationship between collaborative and data-governance aspects. To identify relationships warranting deeper qualitative examination, we applied a light descriptive pre-screening based on the cumulative distribution of importance scores, selecting those at or above the 85th percentile. This threshold is pragmatic, balancing inclusiveness, focus, and analytical feasibility. Relationships below this threshold were not interpreted as unimportant. The cut-off could qualitatively be adjusted case-by-case by reviewing the relevance of relationships both above and below the threshold. A complementary post-hoc weighted analysis was employed to account for variability in participant input density, as some interviewees may have reported only a few relationships while others may have reported many. Therefore, importance scores were recomputed according to the weights and compared with the primary pre-screening results to assess the stability of the outcomes. To ensure minimal empirical support, each selected relationship has to be reported by multiple interviewees and observed in different cases. Therefore, relationships above the percentile threshold, but lacking sufficient qualitative evidence, must be excluded. The outcome of this process is summarized in the *resulting table*, providing a structured representation of the key relationships.

Coding. The relationships among collaborative and DG dimensions were investigated using a structured analytical approach. Initially, the use of *Framework Table* (Appendix A, Fig. A5) facilitated the collection of

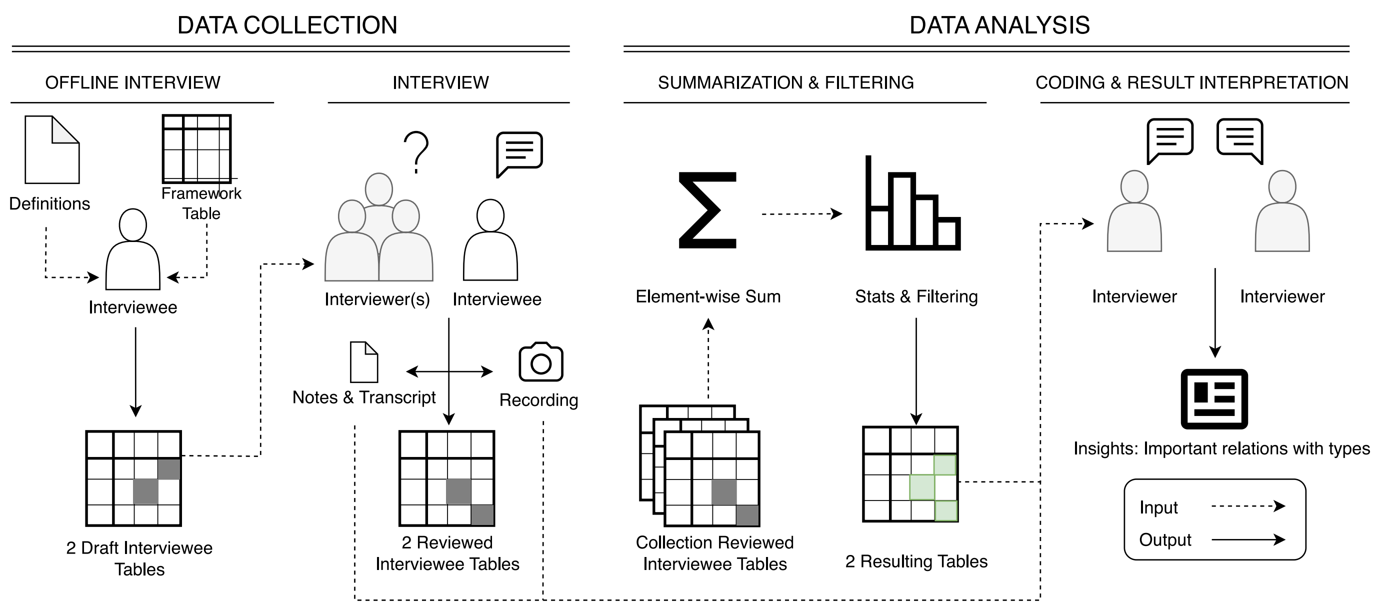

Fig. 2. Flowchart of the research methodology. The elements of the flowchart are referenced in *Italic* in the text.

Table 2
Interviewed participants and their roles across different cases.

Case	Role of Interviewed People
SciExpeM	Software Developer & Platform User
SciExpeM	Client & Platform User
SciExpeM	Client & Platform User
Impact Deal	Design Engineer and Program Manager at operational partner of Impact Deal
Impact Deal	Scientific Director at scientific partner of Impact Deal
E015	Scientific Coordinator of E015
E015	Digital Ecosystem Specialist at E015
Malpensa Smart City	Analyst at SEA Milan Airports
Malpensa Smart City	Software Project Manager at SEA Milan Airports
Malpensa Smart City	Head of Cargo Management at SEA Milan Airports
Health Big Data	Scientific Coordinator of Health Big Data (HBD)
Health Big Data	Quality and Risk Management Director at IRCCS (Partner of HBD)
Health Big Data	Director of Information Systems and Head of Digital Transformation at IRCCS (Partner of HBD)

evidence regarding both unidirectional and bidirectional relationships and ensured the maximum coherence between the interviewees' opinions and the identified relations. Using this systematic approach allowed for the minimization of interviewers' biases in results interpretation and made the process highly replicable. Then, in addition to the two *Resulting Table*, researchers employed a systematic coding process involving multiple iterations among the authors with the aim of identifying the type of relationship. All the interviews were coded manually according to the coding scheme detailed in Appendix C; the coding procedure involved a thorough examination of the data through an in vivo approach. Table A4 in Appendix A presents some examples of specific coded excerpts. To minimize interpretative bias derived from the single author's perspective on the topic, at least two authors coded each interview. The different backgrounds of the authors, alternatively deriving from information systems and organizational studies, helped to strike a balance between different perspectives.

Result Interpretation. After identifying the most significant relationships, we developed an interpretative qualitative analysis, aimed at identifying the conceptual nature of the relationship identified (Rabinovich & Kacen, 2010). This phase aids in the more accurate description of the type of relationship between the governance dimensions, identifying temporal, causal, or contained relationships. Following the framework developed by (Dey, 2003), we distinguished six distinct categories of relationships between variables: concurrent, chronological, overlapping, embedded, explanatory, and causal (see Appendix C). We did so by following the multiple-step process suggested by (Adu, 2019). Starting from the coding of the interview, the process involved the formulation of a sentence describing the relation between the variables and its subsequent classification in the six categories(Dey, 2003). To ascertain the robustness of the analysis, each author initially conducted an independent analysis of the qualitative data derived from the interviews and independently classified each relationship. We followed this step with a discussion to reach a consensus on a unified classification. This additional analysis allowed the unveiling of the conceptual and logical nature behind each core relation identified through the previous research steps, thus enabling the integration of the research findings into practice (i.e., the design and management of governance models).

4. Findings

This section presents the cross-case empirical findings obtained by applying the methodology outlined in Section 3 to all interviews collected from the five cases (Table 1). Across both directions, from CG to DG and vice versa, the interviews yielded a total of 763 reported relationship occurrences. Applying an 85th percentile cut-off (corresponding to scores ≥ 9 and ≥ 8), 28 relationships were initially identified; of these, 24 core relationships were supported by sufficient empirical evidence and retained for detailed analysis. Table 3 presents a concise summary of the core relationships alongside their relationship type, while Fig. 3 visualizes the overall patterns and directionality of the identified links. Extended analytical results and supporting evidence are reported in Appendix B.

The remainder of this section is structured as follows: The first section describes the influence of CG aspects on DG, while the second examines the reverse, focusing on the influence of DG aspects on CG.

4.1. Collaborative governance dimensions' influences

Using a score of 9 as the minimum threshold, 17 relations from CG to DG were found as the most relevant; however, 4 of them were excluded

Table 3
Overview of core relationships and relationships types.

Collaborative Governance to Data Governance	
Description	Relationship Type (s)
Trust enables data transparency	Causal
Trust is a precondition and enables data collection & transformation	Chronological, Causal
Incentives explain data quality	Explanatory
Incentives enable data collection & transformation	Causal
Competences are a precondition to data collection & transformation	Chronological
Competences are a precondition to data quality	Chronological
Flexibility enables data collection & transformation	Causal
Flexibility enables and interrelates with data accessibility & findability	Causal, Overlapping
Organizational structure explains data collection & transformation	Explanatory
Decision making explains data collection & transformation	Explanatory
Community & stakeholder engagement enables data collection & transformation	Causal
Community & stakeholder engagement enables data quality	Causal
Community & stakeholder engagement enables and explains data transparency	Causal, Explanatory

Data Governance to Collaborative Governance	
Description	Relationship Type(s)
Data collection & transformation influence decision making	Causal
Data quality fosters and explains trust	Causal, Explanatory
Data quality influences incentives	Causal
Data quality influences decision making	Causal
Data quality fosters and explains community & stakeholder engagement	Causal, Explanatory
Data accessibility & findability align with decision making	Concurrent
Data privacy aligns, explains, and enables trust	Concurrent, Explanatory, Causal
Data transparency aligns, explains, and enables trust	Concurrent, Explanatory, Causal
Data security aligns, explains, and enables trust	Concurrent, Explanatory, Causal
Data auditing fosters trust	Causal
Data auditing fosters organizational structure	Causal

for lack of empirical evidence (reported in orange in Fig. B6, Appendix B) We describe them below by reporting evidence that emerged from the cases. Each paragraph delineates all of the core relationships that have been identified between a single governance dimension and those that are influenced by it.

Trust The role of trust as a necessary condition influencing data collection and data transparency in DEs has been highlighted by several actors. An interviewee from the Impact Deal case study emphasized the specificity of public sector by highlighting the importance of pre-existing elements of social licensing, which are typically granted to public entities. The same interviewee noted that *"data collection is primarily influenced by social and organizational factors. These factors, in turn, can be shaped by full disclosure of the purposes for which data will be used, thereby increasing trust among the actors involved"*. To sum up, trust enables and is a precondition to data collection with a chronological and causal relationship. Additionally, interviewees underlined the role of trust in ensuring transparent data by accurately processing and transforming the data shared. As remarked by the Software Developer and Platform User of SciExpeM, *"if there is a relationship of trust, people are somehow less reluctant to hide things, even when we are not sure about the quality of what has been done"*. Therefore, it emerges that trust enables data transparency with a causal relationship.

Incentives Another aspect influencing data collection and transformation, as well as data quality, is the presence of incentives. According to an interviewee from the Impact Deal case, market mechanisms and institutional pressure can be utilized to incentivize data collection and transformation. This was observed both in the case of corporate enterprises in the context of Impact Deal. Furthermore, this causal relationship is bidirectional: not only do incentives enable data collection, but frictions or challenges in data collection can act as significant disincentives to participation. Regarding the role of incentives for ensuring data quality, an interviewee from the Malpensa Smart City case noted that low data quality can serve as a warning signal, suggesting the need to enhance incentives. Another interviewee from the same case emphasized that that adopting participatory processes served as an effective incentive by increasing stakeholders' willingness to share high-quality data. Similarly, an interviewee from E015 stated that data quality is not for free and good incentives are fundamental to push people sharing high quality data. Therefore, we can say that incentives explain data quality and identify the relationship type as explanatory.

Competence Competence has emerged as a critical component of CG that affects a variety of DG aspects. As an enabler, the level and type of expertise introduced by the partners affect almost every aspect of DG. In order to maintain specificity, we have chosen only those relationships on which we were able to collect a greater amount of qualitative evidence, such as the relationship with data quality and data collection and transformation. First, the interviewee from the Health Big Data case emphasized that *"general training is necessary; competence must be present*

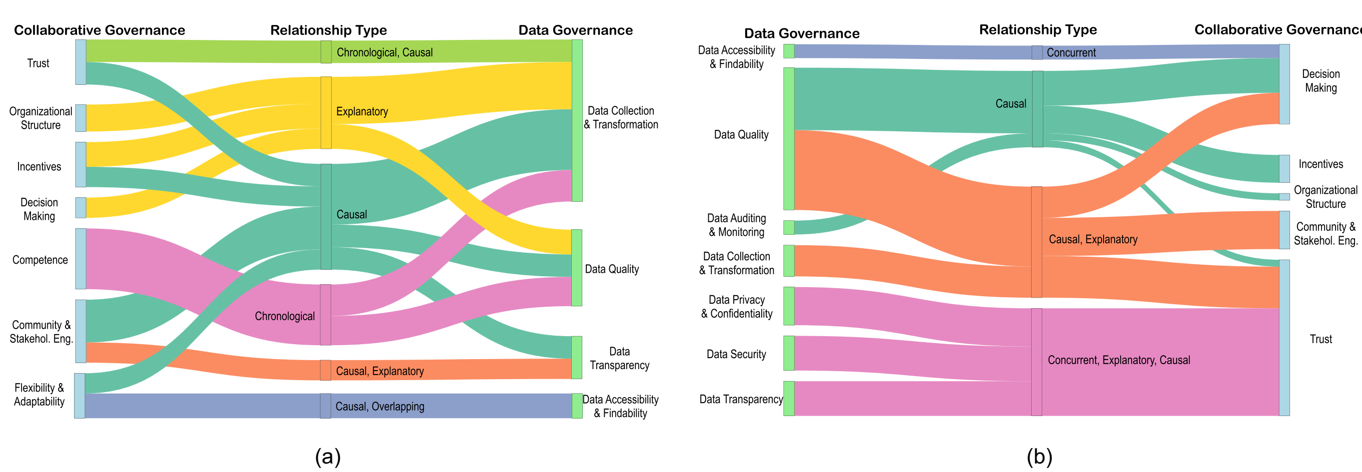

Fig. 3. Sankey diagrams summarizing the 24 key relationships identified in the analysis. Flows run from left to right. Fig. 3a shows relationships from collaborative governance to data governance, while Fig. 3b depicts the reverse direction. The middle column indicates the type of relationship (e.g., causal, chronological, explanatory). Link thickness reflects the relative strength of each relationship, based on the aggregated importance scores shown in Fig. 6.

at all levels, and the more structured and established the companies are, the more important this dissemination process becomes, as it helps overcome greater resistance, which must be addressed through a data culture". A distinction was also made by several interviewees between technical competence in data management and domain-specific expertise. The latter, in the cases of SciExpeM and Impact Deal, was fundamental for both data collection and transformation, as well as for data quality. Additionally, the importance of domain-specific expertise was also highlighted in the Health Big Data case, where one of the interviewees pointed out that it would be better to avoid separating the two types of competence, advocating instead for creating profiles that combine both domain expertise and data skills. Having these competence is a precondition and essential for improving both data collection and quality. This connection has been thus identified as a chronological relationship.

Flexibility Flexibility emerged as a necessary condition to respond to both internal and external stimuli. These stimuli can arise from data regulation, individual partners, or issues related to data quality/accessibility, all of which may require adjustments during the process. In the case of Malpensa Smart City, it was necessary to first evaluate all possible developments of the process as a way to better understand it, which ultimately allowed for more precise data collection, thereby influencing the data collection process itself. In this context, flexibility is an enabler of data collection with a causal relationship. The level of flexibility also enables data accessibility and findability. For instance, interviewees from the E015 case reported that a flexible DE can better accommodate data accessibility requests, allowing access to data to be expanded or restricted in line with continuously evolving needs. In the Health Big Data case, the need for scalable platforms to be used in a multicentric manner influenced data accessibility in terms of who can view or access the data. In other words, flexibility enables and interrelates with data accessibility and findability with causal and overlapping types of relationships. It is worth mentioning the bidirectionality of the relationship between flexibility and accessibility of the data. As claimed by one of the interviewee of Health Big Data *"Greater data accessibility also enables the creation of previously unimagined solutions, thereby providing more flexibility to the entire structure."*.

Organizational Structure Organizational structure is an element of CG that significantly influences DG, particularly in relation to data collection and transformation. In the case of Impact Deal, it was highlighted that the effectiveness of data collection, the quality of the data collected, and their accessibility and findability were typically higher when a proper organizational structure, embedding data collection processes as a key characteristic, was designed from the beginning rather than addressed following the data collection. As emerged from the SciExpeM case, having an organizational structure designed to facilitate the data collection may also imply having a dedicated role (i.e., a data steward) caring about data collection. In these terms, a good organizational structure explains good data collection with a causal link.

Decision Making The decision making process appears to have a direct influence on the data collection process. Indeed, as evidenced in the SciExpeM and Health Big Data cases, having clarity about who is responsible for the data collection and transformation processes, along with clearly defined responsibilities, facilitated a proper data collection and transformation phase. On the contrary, when responsibilities are not clear, it is easy to get stuck in the decision making process, with direct implications on the data collection. Therefore, decision making explains the data collection. In this regard, one of the interviewees of the Health Big Data project reported, *"The absence of hierarchy among the network users implies that they can stop the whole process if they lack the necessary knowledge to make a decision"*. Similarly, an interviewee from the E015 case emphasized the importance of a clear and concrete decision making process, especially when acting in the public sector sphere, for instance by promoting data collection through tenders.

Community and Stakeholder Engagement Community and stakeholder engagement emerged as a central element influencing data collection and transformation, data quality, and transparency across all cases, in a causal way. One interviewee from Impact Deal emphasized that *"the act of collecting data is always an act of power, between those who measure and those who are measured, and therefore, this must be negotiated in a context of reciprocity"*. The same interviewee noted that the importance of stakeholder engagement applies, with a causal effect, equally in public and private contexts, as involving stakeholders fosters a sense of belonging and encourages more active participation, positively impacting and enabling data quality. It is crucial to make stakeholders feel that they are also generating value for the public good. In the Malpensa Smart City case, stakeholder engagement was pivotal in creating a dialogue to.

raise awareness of the importance of data quality. At the same time, the activity of stakeholder engagement enables and explains the transparency and trustworthiness of data and the platform, as noted by one of the interviewees of the Health Big Data case, underlining a causal and explanatory dependency between the two dimensions *"Your capacity to communicate significantly influences your community participation, the frequency with which they contribute, and so forth. They should recognize that the data they provide is a valuable block for improving the decision-making capacity."*.

4.2. Data governance dimensions' influences

Using a score of 8 as the minimum threshold, 11 relations from DG to CG elements were chosen as the most relevant. We discuss them based on the evidence that emerges from the case studies as follows.

Data Collection Data Collection emerged as an important element influencing decision-making. As reported from the Impact Deal case, *"Often, existing frictions and technical problems on the data collection side made the whole process of data exchange really slow, moving partners that were initially enthusiastic about the project to leave it. The technical feasibility of collecting and transforming data is a precondition to the partners' perception of value."*. Additionally, the requirement of managing the data in accordance with specific standards or procedures may necessitate the modification of the decision-making process, underlining a causal relationship type. This is especially true when sensitive data is involved; in fact, ad hoc processes may be necessary to manage them, such as the implementation of ethical checks and processes. For instance, after a few years of operations, the Health Big Data initiative implemented a centralized structure that developed coordination and monitoring activities to improve the integrity of the data. In the E015 case, instead, it emerged that data collection directly informs decision-making by indicating which procedures require further development to better support and promote such a process.

Data Quality Data quality emerges as a key aspect, influencing the levels of trust, the incentives to participation, the community and stakeholder engagement, and the decision-making processes, with a causal relationship. Higher data quality fosters and explains trust, both in the start-up phase and during the ecosystem regime. Having good data quality fuels trust maintenance in the system and encourages stakeholders to share more data and start new collaborations. In this regard, one of the interviewees from SciExpeM reported *"good quality data contributes to building mutual trust, both between those who use the data and collect it and those who manage it"*. Conversely, as emerged in the case of Malpensa Smart City, low-quality data is a strong disincentive for participation in the ecosystem and this is particularly true when a partner is accessing the ecosystem for the first time. At the same time, low data quality can constitute an alarm system signaling the wrong incentive schema, thus requiring its re-definition. This aspect emerged clearly from the Impact Deal case, where one of the interviewees reported, *"How well the data is curated, how well it's maintained, how up-to-date it is, how easy it is to find, how easy it is to export, and making it accessible to third parties are all elements that influence the power of the incentive, the power of the business model that I build, and the return on the investment that I've made in terms of managing the data."*. Similarly, as emerged from multiple cases, good data quality allows for a better understanding of the internal processes of the ecosystem and thus enables their management. Data quality is also subject to critical mass logic; indeed, as emerged from the Health Big Data case, the quality of data and the participants' perception of being key contributors to nurturing and maintaining this quality are reciprocally reinforcing elements. A well-managed communication and engagement strategy may be an important element to activate this virtuous cycle and reach a critical mass effect.

Data Accessibility and Findability Data accessibility constitutes a critical factor in the engagement of stakeholders. In fact, the Impact Deal Case study has demonstrated that data quality and accessibility have a direct impact on the ecosystem's ability to encourage stakeholder participation and demonstrate the return on investment of data sharing. Additionally, as demonstrated by SciExpeM, effective data accessibility facilitates decision-making for both internal (i.e., the ecosystem management) and external stakeholders (e.g., policymakers or data providers), making them more willing to participate. Particularly important in this regard is the accessibility and findability of data, which ensures faster and better decision-making, including the design of ad-hoc incentive systems for ecosystem development. Interestingly, in the E015 case, data accessibility and findability are especially relevant because participating companies can leverage them to increase their visibility and promote their services. Thus, we can affirm that there is a concurrent effect between data accessibility and findability and stakeholder engagement.

Data Security & Privacy & Trasparency Data security and privacy have an impact on the trust levels within ecosystems. It's interesting to note that the data gathered indicates that they both serve at the same time as prerequisites for the ecosystem's activity and enablers of its development. On one side, the ecosystem will not be trusted by any stakeholder unless the technological and organizational configurations that have been implemented provide assurances of security and privacy. Likewise, as stated by one of the interviewees in the Malpensa Cargo case, the assurance that the ecosystem is capable of ensuring specific conditions will enhance trust and encourage participation, particularly when competing companies operate on the same platform, as in the E015 case. In addition, as emerged from the Impact Deal case, the perceived security and privacy level is influenced by the pre-existing trust and social licensing level that stakeholders recognize in the entity responsible for administering the ecosystems. This is particularly true when public entities are involved in data sharing. All these elements define these relationships as circular, concurrent, explanatory, and causal, meaning that the three dimensions mutually reinforce and influence one another.

Data Auditing Data Auditing emerged to foster trust and the organizational structure. More specifically, having in place good data auditing processes may support the smooth operations of the DE, avoiding issues of missed expectations and loss of trust. One of the interviewees from the Impact Deal case reported *"data auditing, especially with respect to partners, should be done a priori so that this cannot have negative repercussions on others in terms of trusts and incentives to participate"*. Additionally, data auditing results in a strong influence on the organizational structure via the enhancement of the capacity to make decisions regarding the ecosystem development. Indeed, as emerged from almost all the cases analyzed, and particularly in the Malpensa Smart City and the Health Big Data case, the analysis of which data are used and how and from whom are all aspects of the data auditing that directly influence the decision making in terms of incentive systems' design and the organizational structure. In this regard, one of the interviewees from the Malpensa Smart City reported, *"Data audits and monitoring enable me to determine where to invest, where to place incentives for which actors, and what themes. Every piece of information I get assists me in determining whether we also need to reshape the ecosystem's organizational structure"*.

5. Discussion

5.1. Discussion of findings

Results indicate the existence of 24 strong relationships between aspects of CG and DG. The identification of these relationships addresses RQ1, highlighting the degree of correlation between the two dimensions, qualifying DE as socio-technical system. This perspective already emerged in various empirical projects (e.g., (The New Hanse Project, 2023), (Otto & Jarke, 2019)) and has been emphasized by a few academic studies (Liva et al., 2023; Otto & Hompel, 2022). This work complemented this perspective by adopting an innovative methodology that allowed us to contemporaneously analyze the existing bidirectional relations among DG and CG aspects, unveiling the intricacy of the governance framework, which has not been made explicit in these terms before. These results imply the necessity of addressing the issue from a renewed holistic perspective entailing the relations between the two dimensions as a key dynamic in the governance of DE.

Adopting an IG approach implies the conjunct conceptualization, design and development of data and CG aspects. This perspective contrasts with the dichotomy of *need-pull* and *technology-push* paradigms in information systems (Chau & Tam, 2000), thereby, advocating for a more integrated approach.

Answering to RQ2, we studied the nature of the relations between organizational and DG dimensions. Results show the different nature of the identified relationships that are alternatively causal, explanatory, concurrent, and chronological or a combination of these. Results enable a more profound analysis of their dependencies and reciprocal effects. For instance, our research reveals a robust causal and explanatory relationship that goes from data quality to trust, suggesting the implications that the former has on the latter. Conversely, results unveil the existence of causal, explanatory and chronological relations between different CG factors (respectively community and stakeholder engagement, incentives and competences) towards data quality. This suggests that in cases of low data quality, interventions should focus on ensuring that those dimensions are well managed within the ecosystem. Although requiring contextualization, these findings are applicable to DE in different context, for instance, the role of competences is particularly critical in the context of public data ecosystems, where there is a need for capacity building in data literacy and digital skills among all stakeholders (Nikiforova et al., 2025). Thus, governments and organizations should invest in training initiatives, such as dedicated programs, workshops, and hackathons, to enable actors to effectively use and manage public data, while having positive effects on data quality. The characteristics of public data ecosystems may also amplify the importance of trust in data collection and transformation, which in our results emerges as both a precondition for and an enabler of data collection, suggesting a chronological and causal relationship. Prior research has similarly underscored the critical role of trust among the various actors involved (Lnenicka et al., 2024). Our results confirm that communicating to stakeholders the benefits they may derive from active participation in the data ecosystem and from sharing their data remains a major concern of both DE and public data ecosystems (Gelhaar & Otto, 2020).

RQ1 and RQ2 allowed us to test and define the concept of **IG** represented in Fig. 4. Through the definition of this concept, we transpose a theoretical discussion to the empirical level, tightening the gap between academic debate and empirical practices. The model does so by describing the relationship types among dimensions, hence unveiling the implications of the socio-technical nature of DE's on their governance. The model answers the call of many to embed a socio-technical point of view into the management of DE (Liva et al., 2023; Micheli et al., 2023; Schmeling et al., 2025). The research resulted in the conceptualization of the Integrated Governance framework displayed in Fig. 4, which highlights the overlap between aspects of CG and DG, and only reports five types of relationships, excluding the embedded

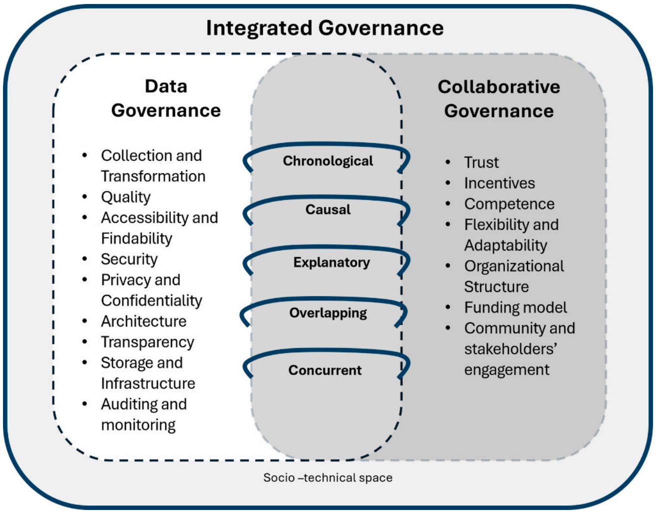

Fig. 4. Integrated governance framework.

relationships type from it, as this type of relationship was not identified in the empirical study.

Results lead to the definition of the concept of Integrated Governance as "the integration, in a unified perspective, of collaborative and data governance aspects and the consideration of their chronological, causal, explanatory, overlapping and concurrent relations as key determinants of the governance model's effectiveness". This definition aligns with the evolution of both the concepts of CG (Ruijer, 2021) and DG (Janssen et al., 2020), which had respectively embedded in recent years technological and organizational elements. However, unlike existing conceptualizations that primarily focus on the constituent elements of governance models, our definition emphasizes the relationships among these elements as a critical determinant of governance effectiveness, thereby advancing a novel line of inquiry.

5.2. Contributions

The first contribution of our study is the creation of a shared vocabulary that facilitates communication between organizational and technical profiles. In this regard, our research answers the call of Schmeling et al. (2025), asking to build semantic ontologies and controlled vocabulary to be used to describe data resources meaningfully. The need for a shared vocabulary was also evident in our experience; indeed, during the interviews conducted for this study, despite sharing the dimension definitions in advance, it became evident that interviewees' diverse backgrounds led to varying interpretations of the same concepts. Not only that, but we found that there is an additional discrepancy between the dimensions of analysis as understood by practitioners and literature. Whilst literature establishes a distinction between governance processes and structures (Agranoff, 2006; Bryson et al., 2015; Emerson et al., 2012), the practitioners' understanding of specific dimensions indicates a partial overlap between certain aspects. For instance, we have observed that practitioners struggle to differentiate between formal decision-making procedures, which are regulated by formal agreements, and the organizational structure. In this regard, our research contributes to establishing a shared vocabulary facilitating cross-disciplinary discussions and bridging the comprehension gap between professionals from different domains and between literature and practitioners, thus facilitating value extraction and co-creation (Mengcheng & Tuure, 2022).

A second theoretical contribution is the reinforcement of the conceptualization of DE as socio-technical systems and its transposition into an integrated and empirically oriented framework. Indeed, although literature has already recognized the socio-technical nature of DE (Fang et al., 2024; Gelhaar et al., 2021; Lnenicka et al., 2024), our study substantiates this perspective by zooming in and framing for the first time the entity and nature of the relationships among social (i.e., collaborative) and technical (i.e., data) dimensions. The definition of the qualitative nature and intrinsic meaning of these relations contributes to theory building on DE governance and complements previous studies affirming the existence of these relations without deepening their nature (Fang et al., 2024; Gelhaar et al., 2021). These evidence invite further reflection on how a socio-technical approach can inform and reshape the design and management of DEs. Moreover, they provide practical insights for practitioners, supporting more informed decision-making and the implementation of corrective actions that can be adopted during both the design and implementation phases of DEs.

Building on the previous ones, another contribution of this study lies in its theory building effort, articulated through the definition of IG as the consolidation of CG and DG. By defining, conceptually and empirically grounding IG, this study advances the data governance literature (Abraham et al., 2019; Donaldson & Walker, 2004; Otto, 2011) as well as the more recent literature on collaborative governance in data-sharing environments (Bartolomucci & Bartalucci, 2024; Micheli et al., 2023; Oliveira et al., 2019; Ruijer, 2021; Susha et al., 2022).

In line with earlier consolidation efforts (Janssen et al., 2020, Ruijer, 2021) and responding directly to recent calls for further conceptual consolidation and theoretical clarification (Möller et al., 2025; Schmeling et al., 2025), this study introduces the concept of IG. IG introduces a new perspective that transcends the cultural and disciplinary silos that have so far limited cross-fertilization between CG and DG and resulted in a fragmented research environment (Schmeling et al., 2025).

5.3. Limitations and future research

Through the framework definition, the research opens up new research venues at the crossroads of information technology and organizational studies, whose fragmentation has so far hampered the development of the sector.

First, given the qualitative and exploratory nature of this study, it does not allow meaningfully testing the strength of the identified relationships, as well as the implications of the adoption of an IG perspective. Thus, further research may investigate the identified relationships more in-depth, particularly in terms of their intensity and medium term implications, while also adapting our findings to different DE forms. Additionally, it should explore how the adoption of an IG perspective may change the design and management of DE.

We also recognize the need to continue to refine the concept of IG by incorporating relationships between the dimensions already considered and the legal and regulatory aspects, which play a key role in defining the boundaries of data sharing (Lazarotto, 2022; Liebert, 2025). We anticipate that these aspects will have reciprocal influences on the CG and data governance elements examined in this study. Complementing the IG model with legal and regulatory integration, as well as developing design processes and tools that facilitate its implementation, represents an important step forward in structuring and managing DE effectively.

In terms of sample selection, our research presents two main limitations. The small sample size of our research may hamper the generalizability of our findings. This limitation is reinforced by the predominant public nature of the cases included in our sample. The combination of the two may have introduced an endogenous sampling bias. Additionally, the research is not exempt from interpretative biases derived from the researchers' stance. The combination of these factors may have influenced our understanding of certain relationships, which may have been thus overlooked or over considered. This is consistent with the literature emphasizing the domain- and sector-specific nature of data ecosystems. Indeed, the environment or setting in which a DE emerges shapes both the types of data that become central within the ecosystem and the concerns of the actors involved (Gelhaar et al., 2021; Oliveira et al., 2019). For example, Oliveira et al. (2019) note that in scientific DEs, actors are primarily concerned with sharing scientific data within academic or research communities. In our study, the funding model was often overlooked by interviewees, which may be attributable to the specific domains represented in our sample. Moreover, sector specific mechanisms may influence the relationships between CG and DG. In sectors like healthcare, the need for a high level of trust and security is paramount (Berlage, Claussen, Geisler, Velasco and Decker, 2022), while in the energy sector, the primary challenge often involves compliance with regulations imposed by data providers (Janev et al., 2022). Thus, we suggest that future studies could replicate the methodology used in this paper in single sectors (e.g., transportation, health) or in environments where specific types of data are used (e.g., open data, sensitive data) for different purposes such as public good, citizen science, etc. also considering how different stakeholders involved in the DE perceive the importance of the relationships identified. Additionally, results' robustness may also benefit from their verification through different research methodologies, like in depth case studies or focus groups that may also test it against the authors' interpretative bias.

Another potential avenue for future research would be applying the model developed in this study to DEs in a specific stage of their lifecycle (e.g., initiation, growth, maturity) or studying their evolution through time. Indeed, as emerged in similar contexts (Bartolomucci & Leoni, 2024; Ooms, Caniëls, Roijakkers, & Cobben, 2020), the governance of collaborative ecosystems evolves along the ecosystem's life cycle, which may lead to the predominance of certain relationships over others at different stages.

6. Conclusions

Addressing the lack of empirical evidence on the relationship between organizational and technical factors in DEs, this study aims at investigating what the most relevant influences are between CG and DG factors within DEs and what the nature of the core relationships between these two dimensions is. Adopting a multiple case studies approach, we investigate relationships between CG and DG across five DEs operating in diverse sectors, including health, logistics, and research.

Through a structured multi-step interpretative qualitative analysis, the study provides insights on 24 key relation- ships between CG and DG, highlighting the different implications these relationships have for both theory and practice. Each relationship identified is alternatively categorized as causal, chronological, explanatory, overlapping, concurrent, or a combination of these. From a theoretical perspective, the study introduces a conceptual framework of *Integrated Governance (IG)*, providing a novel perspective on how to approach DE governance, contributing to the ongoing academic discussions. From a practical perspective, the findings of the study elucidate the relationship between CG and DG mechanisms. Practitioners involved in designing and managing collaborative environments, by understanding the interplay between the governance dimensions, can better align their strategies and avoid potential misalignment. By positioning these insights within broader discussions on socio-technical systems and policy coordination, the study establishes a foundation for future research to extend and adapt these frameworks across diverse organizational contexts and data-sharing environments. Overall, our result represents a preliminary exploration that can be further deepened in future research. The interpretative analysis facilitates additional theory development regarding the relationships among the identified dimensions and creates room for quantitative research aimed at empirically validating these relationships across different contexts.

Funding statement

No funding has been received to develop this research.

CRediT authorship contribution statement

Federico Bartolomucci: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Edoardo Ramalli:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization, Validation. **Valeria Maria Urbano:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Appendix: Definitions

This section provides detailed definitions of the collaborative and data-governance dimensions. Fig. A5 shows the *Framework Table* thar interviewees were asked to fill in twice. Table A4 illustrates the coding process with specific excerpts, showing how raw qualitative input was translated into relationship codes for each relationship type.

Data Governance

Collaborative Governance	Collection & Transformation	Quality	Accessibility & Findability	Security	Privacy & Confidentiality	Modeling	Transparency	Storage & Infrastructure	Auditing & Monitoring
Trust									
Incentives	+1								
Competence			+1						
Flexibility & Adaptivity									
Organizational Structure					+2				
Decision Making									
Funding Model									
Community Engagement									

Fig. A5. *Framework table*: The figure displays the *Framework Table* that participants were asked to complete with a few examples. The rows are the dimensions of CG, while the columns are the DG dimensions. Interviewed participants mark the presence of a relationship with a value of 1, and its absence with 0 (zeros are omitted for readability). The most critical relationships are highlighted with a value of 2.

A.1. Collaborative governance aspects

Trust (TR) is a complex and dynamic concept that evolves over the course of collaboration and that can be influenced by intentional trust-building efforts (Bryson et al., 2015). Trust is a critical factor in the establishment of shared motivation to achieve the collaborative objective (Mengcheng & Tuure, 2022), as it fosters reciprocal understanding, legitimacy, and commitment (Emerson et al., 2012). The potential catalyst effect between trust and successful collaborative activities (Klievink, Van Der Voort, & Veeneman, 2018), suggests its relationship with data governance aspects. Beyond the conventional trust dynamics in collaborative ecosystems, trust in data ecosystems is inherently linked to data- specific activities, particularly when sensitive information is involved, suggesting a relation with data security, protection and privacy dimensions (Haak, Ubacht, Van Den Homberg, Cunningham, & Van Den Walle, 2018; Urbano, Bartolomucci, & Azzone, 2024). Additionally, robust data infrastructures and governance processes are essential (Stalla-Bourdillon, Wintour, & Carmichael, 2019), alongside continuous communication and engagement cycles (Farmer, McCosker, Albury, & Aryani, 2023).

Incentives (IN) extend beyond simple financial rewards. They encompass the positive and negative changes in outcomes that individuals anticipate from their actions within a specific set of rules, influenced by physical and social contexts (Ostrom, 2009). Regulatory incentives serve as a key example of non-monetary incentives. As policy instruments and integral components of governance systems, they can be used to directly or indirectly influence behaviors, hence promoting desired outcomes (Ponti et al., 2024). Incentives are contingent on actors' expectations and may vary significantly based on the nature of the ecosystem, be it altruistic or market-driven, and on the nature of the actors involved. While public and non-profit organizations may be moved by the aim to generate positive impact, private actors are mostly incentivized by economic and efficiency gains (Klievink et al., 2018). The possibility to gain knowledge and insight over data and increase brand equity and legitimacy are the major incentives found in data sharing environments (Moretti, Zahuranec, & Verhulst, 2022). Still, incentives have to be considered as part of a risk-benefit analysis, especially in relation to the time and effort required for collaboration (Ansell & Gash, 2008). Framing it under the privacy-calculus theory (Jussen, Schweihoff, & Moller, 2023) identifies the interplay of organizational, business model, and data sovereignty factors in the creation of multiple tensions in data sharing, among which are those connected to data usage, data control, and data trust.

Competence (CO) refers to the combination of competence that each participant possesses, as well as the collective intelligence generated by all actors participating in the collaborative environment. The knowledge mix and distribution among the actors may influence the spectrum of feasible activities in the data ecosystem, as well as generate power imbalances between the actors. Matching the combination of knowledge, competence, and capacity to the intended interventions at its outset can lower the communication load among actors, improve reciprocal comprehension, and limit the risk of blockers. In this vein, literature has recently highlighted the concept of Data Stewards (van Donge, Bharosa, & J. M., 2022) as a potential critical resource to include in enterprises participating in data ecosystems. This role may enable optimal inter- and intra-organizational handling of data, resulting in benefits such as higher data quality, accessibility, and interoperability.

Flexibility and Adaptability (FA) pertain to the capacity of the data ecosystem to be adaptive to internal and external pressures for change. Pressure for change may come from internal aspects of the data ecosystems, such as tensions among stakeholders or dissatisfaction given by the effort needed to reach certain objectives. Alternatively, it may arise from external factors like new regulations. In both cases, the trade-off between organizational flexibility and the degree of determinism of scopes and roles is a key element to address (Provan & Kenis, 2008). Indeed, the latter may alleviate the collaborative burden at the outset, but it may also increase the partnership's susceptibility to crises in the long term. This organizational characteristic may be significantly influenced by the data modeling, infrastructure, and data findability dimension's (Otto, 2011).

Organizational Structure (OS) The concept of organizational structure in collaborative ecosystems entails all those elements that refer to formal structures and formalized processes that regulate the ecosystem's functioning (Bryson et al., 2015). The organizational structure may be more or less centralized. Less centralized structures are those in which responsibilities, decision power and technological infrastructures are distributed among the actors. Centralized structures are instead those in which one single actor plays the role of intermediary (Janssen & Singh, 2022). In most advanced forms, intermediation involves decision-making, including on behalf of the partner, revenue generation (Susha et al., 2022), data stewardship (Flanagan Anne & Sheila, 2022), as well as infrastructure provision. Literature recognizes the role that intermediaries can have in building trust, solving conflicts and disintermediating relations among the partners, as well as managing data accessibility and use (Flanagan Anne & Sheila, 2022) and ensuring data quality (van Donge et al., 2022).

Decision Making (DE) refers to the informal collaborative processes that engage partners in the process of decision- making, among which those over data-related issues. The decision making process is linked to communication management, which can directly affect collaborative processes (Koschmann, Kuhn, & Pfarrer, 2012). Formal decision-making derives from the organizational structure adopted; the informal one does not, and it's more related to power dynamics. Both can influence data processes. For instance, the presence or absence of an ethical committee is one of the organizational dimensions affecting decision-making processes. The decision-making process also influences the partnership's flexibility and the level of trust among the involved parties. Topics related to data accessibility, data privacy and confidentiality may also be related to this dimension.

Funding Model (FM) refers to the way in which data ecosystems sustain their activity. This dimension, often excluded from governance frameworks, has been recently reconsidered as part of it (Carballa Smichowski, 2019; Susha et al., 2020). Literature on data partnerships frequently identifies it as a key governance challenge and emphasizes its significance for partnership stability (GSMA, 2018). (van Donge et al., 2022) identifies the co-development of responsible business models as one of the governance model design's outputs. (Susha et al., 2020) identifies four distinct business models for data ecosystem intermediaries, with each of them providing distinctive value propositions and varying in terms of data ownership and control, bargaining power in decision-making, and thus degree of reliance on other network participants.

Community & stakeholders' engagement (CE) The aspect of community interaction focuses on identifying and understanding the dynamics of interaction between the data ecosystem's managers and the communities it serves (Sarker et al., 2021). Stakeholders' engagement is known to be an effective way of gaining trust and legitimacy as well as being able to better fit solutions to problems (Mainka et al., 2016). Including stakeholders and communities may increase their sense of engagement and participation, with direct positive effects on data quality, relevance, and richness (Haak et al., 2018). This is particularly true in contexts in which stakeholders are the primary data owners and sources, like in data cooperative or data crowdsourced environments (Williams, 2020).

A.2. Data governance aspects

Data Collection & Transformation (DC) Data collection involves gathering raw data from various sources (Sapsford & Jupp, 1996). It is crucial to ensure that data collection methods are reliable, accurate, and compliant with legal and ethical standards, especially regarding privacy and consent. Data transformation refers to all aspects and activities that make raw data into a usable form for analysis and storage (Famili, Shen, Weber, & Simoudis, 1997). Among these data pre-processing tasks, there is data cleaning (e.g., removing errors, duplicates, or inconsistencies), data integration (combining data from multiple sources), data normalization (ensuring consistency in format and units), and data enrichment (adding additional information or context to the data). Without the proper data transformation activities, making the data usable for downstream applications will not be possible.

Data Quality (DQ) The data quality assessment refers to the evaluation of data to determine whether they retain a sufficient grade to be used (Pipino, Lee, & Wang, 2002). Over the years, many data quality dimensions have been defined, such as accuracy, timeliness, and completeness, but these dimensions are relative and need to be defined in the context of the end use of data. It has been demonstrated that low quality in the data can have huge repercussions in the applications with high loss in revenues (Batini, Scannapieco, et al., 2016). Similarly, the quality of the data directly affects the incentives to use and share data within a platform (Oliveira & Lóscio, 2018).

Data Accessibility & Findability (DF) According to the FAIR principles (Wilkinson et al., 2016), data accessibility refers to the ease with which data can be retrieved, accessed, and utilized by users, ensuring that appropriate mechanisms, formats, and permissions are in place. On the other hand, data findability is related to the ability to locate and discover relevant data for a user efficiently. Strategies such as structured metadata, proper indexing, and search functionalities are central to achieving such scope. Ensuring that the platform's data is both accessible and findable enhances user engagement, fosters trust and encourages stakeholders to contribute, share and reuse data more effectively.

Data Security (DS) Data governance aspects related to data security include defining data security policies, standards, and procedures to guarantee authentication (i.e., user identification), authorization (i.e., granting the users the proper permissions), and integrity (i.e., data is accurate, real, and safeguarded) of the data during the data life cycle (Singh, 2020). For instance, a data ecosystem without the proper security measures can jeopardize the trust in the platform itself, disincentivizing users to provide and consume data since it can be tampered with or be of ambiguous provenance.

Data Privacy & Confidentiality (DP) Data privacy and confidentiality are related to the actions to safeguard sensitive or private information from unauthorized access or alteration. It implicates implementing ad-hoc measures and protocols (Donaldson & Walker, 2004; Matthews & Harel, 2011). Governance decisions regarding data pri- vacy and confidentiality are related to the data security measures in place but go beyond that. They define the scope of data sharing, establish policies for access control, and determine which mechanisms guarantee that privacy and confidentiality are ensured while maintaining the usability of the data platform. Such mechanisms may strongly influence trust-building in the platform users.

Data Modeling (DM) Data modeling regards the design, implementation, and maintenance of data models that are the foundations of every data-based application (Castro, Villagra, Garcia, Rivera, & Toledo, 2021). It involves defining data objects and their interconnection by developing conceptual, logical, and physical data models. Proper data modeling is fundamental and should be done prior to the actual development of an application since it is the basis of the entire business logic. It establishes policies, standards, and guidelines, thus influencing decision making. Later modifications to the data modeling are possible but can be disruptive.

Data Transparency (DT) Data Transparency refers to making data and the processes surrounding its collection, usage, and management easily understandable and accessible to stakeholders. The governance of data transparency defines the principles for making every step of the data lifecycle recorded, replicable, and understandable (Geisler et al., 2021). This involves providing, for instance, clear documentation on metadata, including provenance, and explanations regarding the data sources and operations done on the data. To achieve transparency, initiatives such as open data are fundamental. Transparency brings trust among stakeholders, enables informed decision-making, and promotes accountability. At the same time, it allows the delivery of a better end product by allowing different people to validate the data, the process, and the outcomes.

Data storage and infrastructure (DI) To host any data-sharing initiatives, it is necessary to make some decisions regarding the infrastructure, mainly regarding the computational and storage capabilities (Castro et al., 2021). Every infrastructure has non-negligible costs in terms of hardware and software requirements; thus, it needs to be discussed who will finance and maintain such an IT system. Other critical decisions can be regarded as to where to host such infrastructure. For instance, a shared and central computational facility can be cost-effective, while more decentralized structures can guarantee privacy and power distribution. Further governance mechanisms involve defining standards and protocols to make the IT system interoperable with other infrastructures and applications.

Data Auditing and Monitoring (DA) The primary goal of data auditing is to verify that data conforms to established standards, policies, and regulations. It includes tracking data changes, identifying discrepancies or anomalies, and investigating any irregularities that may indicate errors, fraud, or unauthorized access. Data monitoring involves continuously observing and tracking data flows, access, and usage within an organization's systems and networks. It aims to detect and respond to suspicious or unauthorized activities in real time to prevent security breaches, data leaks, or compliance violations. A good data audit could influence trust dynamics.

Appendix B. Appendix: Analytical findings

The analysis of the reviewed interviewee tables reveals that, on average, each interviewee reported 28 relationships from collaborative aspects to data aspects and 23 relationships in the opposite direction, with the number of relationships reported from each interviewee ranging from a maximum of 44 to a minimum of 8, respectively. The cumulative distribution of the score of importance is shown for both resulting tables in Fig. B7. In the cumulative distribution, approximately 20% of the values fall below a threshold of 3. On the other hand, the 50th percentile indicates that a relationship is reported at least five times in both cases. The 85th percentile corresponds to a threshold value of around 9 and 8. In Fig. B6, only the relationships that have higher importance score than these identified thresholds are highlighted, with different shades of green indicating their level of importance, with darker shades representing the most important relationships. The analysis of the resulting tables reveals 414 occurrences of unidirectional relationships from CG dimensions to DG dimensions, while 349 occurrences have been found in the opposite direction, from data governance dimensions to collaborative governance dimensions (Table 3). Given the two-step methodology, which includes the threshold defined in the cumulative distribution analysis (85*th* percentile) and a qualitative evaluation of the relationships through interview coding. Following the quantitative evaluation, we conducted a qualitative evaluation of each dimension with a score greater than the defined thresholds. This second step of the analysis led to the exclusion of 4 dimensions (reported in orange in Fig. B6a) on which, despite receiving high quantitative scores, we were lacking empirical evidence to support them. The complementary post-hoc weighted analysis reinforced the results of the two-steps methodology (Fig. B8), resulting in a total number of 24 core relationships to be considered. This approach allowed us to account for variability in the relative importance of relationships between dimensions, as some participants identified only a few connections, while others identified many. By recalibrating the analysis with these weights and comparing the outcomes with those obtained through the primary two-step methodology, we evaluated the consistency and reliability of our findings.

Table A4
Examples of quotes from the interviews For each relationship type, an example between the collaborative and data-governance aspects is reported, accompanied by an interview excerpt demonstrating how raw qualitative input was translated into a relationship code.

Relationship Type	Collaborative and Data Governance Aspect	Quote
causal	Trust Building – Data Transparency	If both parties believe they will benefit from building the structure, data collection is more actively pursued.
causal & explanatory	Data Quality – Community and Stakeholders Engagement	How data is curated, maintained, updated, easy to find, export, and make accessible to third parties—all these elements influence the development of a business model and the return on investment in data management.
causal & overlapping	Flexibility – Data Accessibility	The possibility of using scalable platforms in a multi-center way influences architecture, infrastructure, and data accessibility. Knowing who can see and access certain types of data allowed us to make the most of what was implemented.
explanatory	Organizational Structure – Data Collection and Transformation	Data collection, the quality of how it's collected, and whether the data is accessible and searchable is typically more effective when implemented by design, rather than added later.
chronological	Competence – Data Collection and Transformation	Technical and domain expertise (in how the information system works) is important to enable the collection and management of data.
chronological & causal	Trust – Data Collection and Transformation	Trust is a necessary condition for initiating data collection. It is influenced by technological factors, but also by social and organizational ones. These can be shaped by full disclosure about the purpose of data use.
concurrent	Data Privacy – Trust Building	In this case, for example, privacy, etc.—I thought that if the user sees that certain privacy and security standards are respected, it could help build greater trust.
concurrent & explanatory & causal	Data Security – Trust Building	A government has a social license that allows it to enforce things, but there is a constant fear of function drift: saying you're providing function X but using it for other purposes. Using security standards and cutting-edge technology can help build trust.

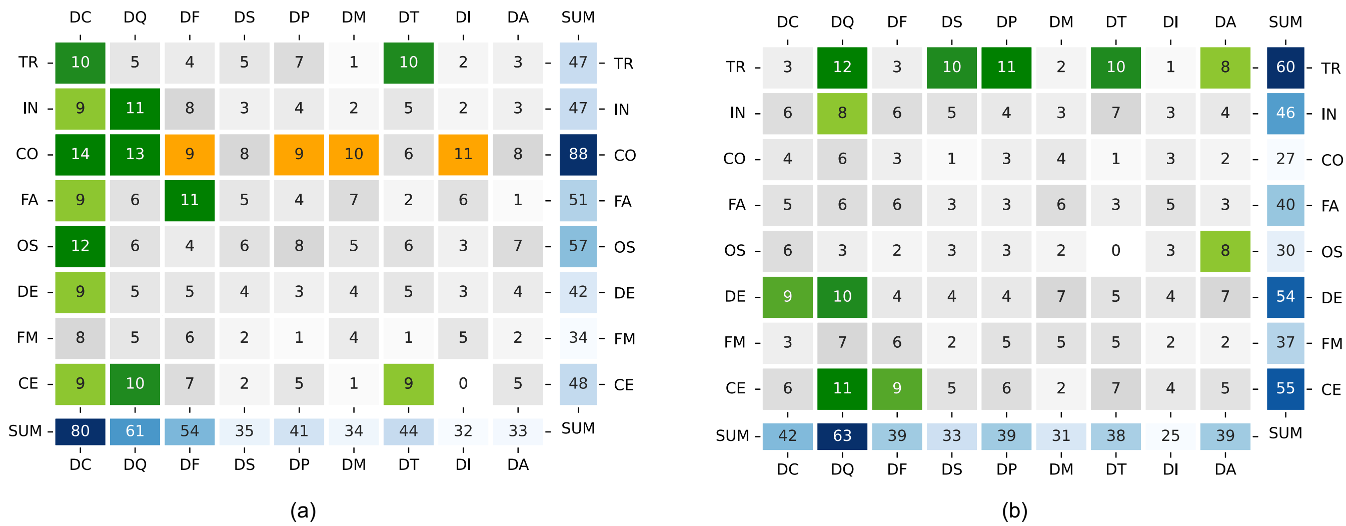

Fig. B6. Resulting tables. Fig. B6a shows the score of the relationships between CG dimensions and data governance dimensions. Fig. B6b shows the score of the relationships between data governance dimensions and CG dimensions. The additional row and column, shaded in blue, display the cumulative sum for each dimension in each resulting table. Darker shades indicate the most relevant dimensions for IG, while lighter shades correspond to those that were less frequently reported as being related to other dimensions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

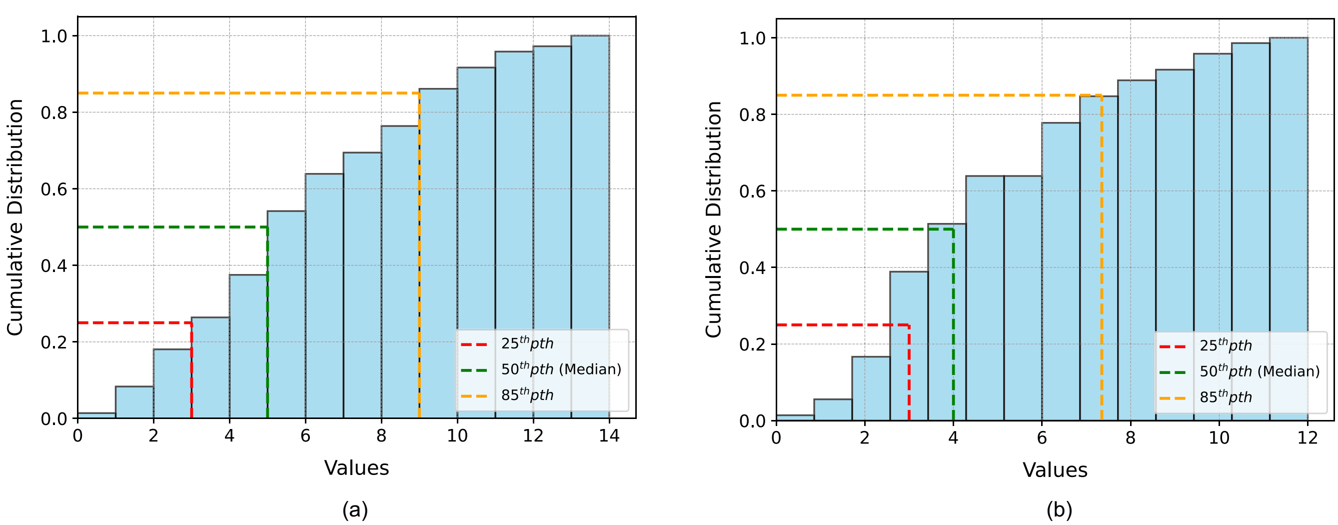

Fig. B7. Cumulative distribution plot of the importance scores. The x-axis reports the score of importance, while the y-axis shows the cumulative proportion of respondents who rated the dimension at or below that score. Fig. B7a presents the distribution for collaborative-governance dimensions, whereas Fig. B7b shows the corresponding distribution for data- governance dimensions.

(a)

(a)

Fig. B8. Weighted resulting tables. Fig. B8a shows the weighted score of the relationships between CG dimensions and data governance dimensions. Fig. B8b shows the weighted score of the relationships between data governance dimensions and CG dimensions.

Appendix C. Appendix: Relationship types

This appendix reports the types of relationships identified by (Dey, 2003) and described by (Adu, 2019). We used these types of relationships and definitions as part of the Research Framework and to interpret our findings. The results are displayed in Table 3.

Concurrent relationship. Two concepts, events, behaviors or processes have a concurrent relationship if they exist, happen, change or impact at the same time.

Chronological relationship. Two concepts, events, behaviors or processes have a chronological relationship if one concept precedes or follows the other.

Overlapping relationship. Two concepts have an overlapping relationship if they have aspect(s) of their characteristics in common in terms of explaining a phenomenon, influencing another concept(s), or representing a concept, process, behavior or an event.

Embedded relationship. Two concepts have an embedded relationship if the entire characteristics of one concept are completely shared with a portion of another concept's features in terms of explaining a phenomenon, influencing another concept(s), or representing a process, behavior or an event.

Explanatory relationship. Two concepts have an explanatory relationship if one concept plays the role of clarifying, elaborating or exemplifying another.

Causal relationship. Two concepts have a causal relationship if the existence of or changes in a concept leads (or contributes) to the emergence or adjustment of another.

References

Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*, 424–438.

Adu, P. (2019). *A step- by- step guide to qualitative Data coding*. New York, NY: Routledge.

Agranoff, R. (2006). Inside collaborative networks: Ten lessons for public managers. *Public Administration Review, 66*(Suppl. 1), 56–65.

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision Systems, 25*(sup1), 64–75.

Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory, 18*(4), 543–571.

Baker, S., Waterfield, J., & Bartlam, B. (2018). Saturation in qualitative research : Exploring its conceptualization and operationalization. *Quality & Quantity, 52*(4), 1893–1907.

Bartolomucci, F., & Bartalucci, S. (2024). Designing data collaboratives' governance dimensions for long-term stability: An empirical analysis. *Data & Policy, 6*.

Bartolomucci, F., & Leoni, F. (2024). Designing an effective governance model for Data Collaboratives. *Research-Technology Management, 67*(4), 49–61.

Batini, C., Scannapieco, M., et al. (2016). *Data and information quality. Cham, Switzerland: Springer International Publishing.* page 63.

Berlage, T., Claussen, C., Geisler, S., Velasco, C. A., & Decker, S. (2022). *Medical data spaces in healthcare data ecosystems*.

Bryson, J. M., Crosby, B. C., & Stone, M. M. (2006). The design and implementation of cross-sector collaborations: Propositions from the literature. *Public Administration Review, 66*(Suppl. 1), 44–55.

Bryson, J. M., Crosby, B. C., & Stone, M. M. (2015). Designing and implementing cross-sector collaborations: Needed and challenging. *Public Administration Review, 75*, 647–663.

Bühler, M. M., Calzada, I., Cane, I., Jelinek, T., Kapoor, A., Mannan, M., … Zhu, J. (2023). Unlocking the power of digital commons: Data cooperatives as a pathway for data sovereign, innovative and equitable digital communities. *Digital, 3*(3), 146–171.

Cappiello, C., Gal, A., Jarke, M., & Rehof, J. (2020). Data ecosystems: Sovereign data exchange among organizations (dagstuhl seminar 19391). *Dagstuhl Reports, 9*(9), 66–134.

Carballa Smichowski, B. (2019). Alternative data governance models: Moving beyond one-size-fits-all solutions. *Intereconomics, 54*(4), 222–227.

Castro, A., Villagra, V. A., Garcia, P., Rivera, D., & Toledo, D. (2021). An ontological-based model to data governance for big data. *IEEE Access, 9*, 109943–109959.

Chau, P. Y., & Tam, K. Y. (2000). Organizational adoption of open systems: A 'technology-push, need-pull'perspective. *Information & Management, 37*(5), 229–239.

Curry, E., & Sheth, A. (2018). Next-generation smart environments: From system of systems to data ecosystems. *IEEE Intelligent Systems, 33*(3), 69–76.

Degen, K., Lutzens, R., Beschorner, P., & Lucke, U. (2025). Public education data at the crossroads of public and private value creation: Orchestration tensions and stakeholder visions in Germany's emerging national digital education ecosystem. *Electronic Markets, 35*(1), 19.

Demchenko, Y., De Laat, C., & Membrey, P. (2014). Defining architecture components of the big data ecosystem. In *In 2014 international conference on collaboration technologies and systems (CTS)* (pp. 104–112). IEEE.

Dey, I. (2003). *Qualitative data analysis: A user friendly guide for social scientists*. Routledge.

Donaldson, A., & Walker, P. (2004). Information governance—a view from the nhs. *International Journal of Medical Informatics, 73*(3), 281–284.

van Donge, W., Bharosa, N., & J. M.. (2022). Data-driven government: Cross-case comparison of data stewardship in data ecosystems. *Government Information Quarterly, 39*, Article 101642.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550.

Emerson, K., Nabatchi, T., & Balogh, S. (2012). An integrative framework for collaborative governance. *Journal of Public Administration Research and Theory, 22*(1), 1–29.
Famili, A., Shen, W.-M., Weber, R., & Simoudis, E. (1997). Data preprocessing and intelligent data analysis. *Intelligent Data Analysis, 1*(1), 3–23.
Fang, J., Zhao, L., & Li, S. (2024). Exploring open government data ecosystems across data, information, and business. *Government Information Quarterly, 41*(2), Article 101934.
Farmer, J., McCosker, A., Albury, K., & Aryani, A. (2023). *Data for social good: Non-profit sector data projects*. Springer Nature.
Flanagan Anne, J., & Sheila, W. (2022). *Advancing digital agency : The power of data intermediaries*. World Economic Forum: Technical Report February.
Geisler, S., Vidal, M.-E., Cappiello, C., Lóscio, B. F., Gal, A., Jarke, M., , … Paja, E., et al. (2021). Knowledge-driven data ecosystems toward data transparency. *ACM Journal of Data and Information Quality (JDIQ), 14*(1), 1–12.
Gelhaar, J., Groß, T., & Otto, B. (2021). A taxonomy for data ecosystems. In *Proceedings of the 54th Hawaii international conference on system sciences (2021)*. Hawaii International Conference on System Sciences.
Gelhaar, J., & Otto, B. (2020). Challenges in the emergence of data ecosystems. In *, 175. PACIS*.
Gil-Garcia, J. R., & Sayogo, D. S. (2016). Government inter-organizational information sharing initiatives: Understanding the main determinants of success. *Government Information Quarterly, 33*(3), 572–582.
GSMA. (2018). *Scaling big Data for social good: The need for sustainable business models. (September)*.
Haak, E., Ubacht, J., Van Den Homberg, M., Cunningham, S., & Van Den Walle, B. (2018). A framework for strengthening data ecosystems to serve humanitarian purposes. In *ACM International Conference Proceeding Series*.
Janev, V., Vidal, M.-E., Pujić, D., Popadić, D., Iglesias, E., Sakor, A., & Čampa, A. (2022). Responsible knowledge management in energy data ecosystems. *Energies, 15*(11), 3973.
Janssen, H., & Singh, J. (2022). Data intermediary. *Internet Policy Review, 11*(1), 1–9.
Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly, 37*(3), Article 101493.
Jussen, I., Schweihoff, J., Dahms, V., Möller, F., & Otto, B. (2023). Data sharing fundamentals: Definition and characteristics. In *HICSS* (pp. 3685–3694).
Jussen, I., Schweihoff, J., & Moller, F. (2023). Tensions in inter-organizational data sharing: Findings from literature and practice. In *, 2023. Proceedings- 2023 IEEE 25th conference on business informatics*. CBI (June).
Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152.
Klievink, B., Van Der Voort, H., & Veeneman, W. (2018). Creating value through data collaboratives: Balancing innovation and control. *Information Polity, 23*(4), 379–397.
Koschmann, M. A., Kuhn, T. R., & Pfarrer, M. D. (2012). A communicative framework of value in cross-sector partnerships author (s): Matthew A, Koschmann, Timothy R. Kuhn and Michael D. Pfarrer Published by: Academy of Management Stable URL : http://www.jstor.com/stable/23218092 A communicative framework. *Academy of Management Review, 37*(3), 332–354.
Ladley, J. (2019). *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press.
Lazarotto, B. (2022). The Data Act: empty promises for business-to-government data sharing? *PinG Privacy in Germany, 5*.
Liebert, P. (2025). On the interplay of governance mechanisms in Data space-enabled ecosystems. *Thirty-first Americas Conference on Information Systems*, 1–10.
Lis, D., Gelhaar, J., & Otto, B. (2023). Data strategy and policies : The role of data governance in data ecosystems. *Intelligent Technologies and Robotics Intelligent Technologies and Robotics, R0*, 27–55.
Lis, D., & Otto, B. (2020). *Data governance in data ecosystems–insights from organizations*.
Liva, G., Micheli, M., Schade, S., Kotsev, A., Gori, M., & Codagnone, C. (2023). City data ecosystems between theory and practice: A qualitative exploratory study in seven European cities. *Data & Policy, 5*.
Lnenicka, M., Nikiforova, A., Clarinval, A., Luterek, M., Rudmark, D., Neumaier, S., … Bolívar, M. P. R. (2024). Sustainable open data ecosystems in smart cities: A platform theory-based analysis of 19 european cities. *Cities, 148*, Article 104851.
Mainka, A., Castelnovo, W., Miettinen, V., Bech-Petersen, S., Hartmann, S., & Stock, W. G. (2016). Open innovation in smart cities: Civic participation and co-creation of public services. *Proceedings of the Association for Information Science and Technology, 53*(1), 1–5.
Matthews, G. J., & Harel, O. (2011). *Data confidentiality: A review of methods for statistical disclosure limitation and methods for assessing privacy*.
Mengcheng, L., & Tuure, T. (2022). Information technology–supported value co-creation and co-destruction via social interaction and resource integration in service systems. *Journal of Strategic Information Systems, 31*(2), Article 101719.
Micheli, M., Farrell, E., Carballa Smichowski, B., Posada-Sànchez, M., Signorelli, S., & Vespe, M. (2023). *Mapping the landscape of data intermediaries. Technical report*. European Commission Joint Research Centre.
Möller, F., Legner, C., Strobel, G., Schoormann, T., Cappiello, C., Loscio, B., & Otto, B. (2025). *Data ecosystems in IS research: The road so far, where we are now, and the road ahead* (pp. 1–8).
Moretti, L. T., Zahuranec, A., & Verhulst, S. G. (2022). *The 9Rs Framework*.
Nikiforova, A., Lnenicka, M., Luterek, M., Milic, P., & Rodríguez, M. P. B. (2025). Theorizing the evolution of public data ecosystems: An empirically grounded multi-generational model and future research agenda. *Government Information Quarterly, 42*(3), Article 102062.
Oliveira, M. I., Barros Lima, G. D. F., & Farias Lóscio, B. (2019). *Investigations into Data ecosystems: A systematic mapping study, volume 61*. London: Springer.
Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem?. In *Proceedings of the 19th annual international conference on digital government research: Governance in the Data age* (pp. 1–9).
Ooms, W., Caniëls, M. C., Roijakkers, N., & Cobben, D. (2020). Ecosystems for smart cities: Tracing the evolution of governance structures in a dutch smart city initiative. *International Entrepreneurship and Management Journal, 16*(4), 1225–1258.
Ostrom, E. (2009). *Understanding institutional diversity*. Princeton University Press.
Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems, 29*(1), 3.
Otto, B., & Hompel, M. (2022). *Designing data spaces*.
Otto, B., Hompel, M. T., & Wrobel, S. (2019). International data spaces: Reference architecture for the digitization of industries. *Digital Transformation*, 109–128.
Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the international data spaces case. *Electronic Markets, 29*(4), 561–580.
Otto, B., Lohmann, S., Auer, S., Brost, G., Cirullies, J., Eitel, A., Ernst, T., Haas, C., Huber, M., Jung, C., et al. (2017). *Reference architecture model for the industrial data space*.
Panian, Z. (2010). Some practical experiences in data governance. *World Academy of Science, Engineering and Technology, 62*(1), 939–946.
Pardo, T. A., Gil-Garcia, J. R., & Luna-Reyes, L. F. (2010). Collaborative governance and cross-boundary information sharing: Envisioning a networked and it-enabled public administration. In *The future of public administration around the world: The Minnowbrook perspective* (pp. 129–139).
Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218.
Ponti, M., Portela, M., Pierri, P., Daly, A., Milan, S., Lindholm, R. K., … Gonzalez, S. T. (2024). *Unlocking green deal data: Innovative approaches for data governance and sharing in europe*. Technical Report, European Commission, Joint Research Centre (JRC).
Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory, 18*(2), 229–252.
Rabinovich, M., & Kacen, L. (2010). Advanced relationships between categories analysis as a qualitative research tool. *Journal of Clinical Psychology, 66*(4), 430–441.
Ramalli, E., & Pernici, B. (2023). Sustainability and governance of data ecosystems. In *2023 IEEE international conference on web services (ICWS)* (pp. 740–745).
Ruijer, E. (2021). Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly, June*, Article 101612.
Salnitri, M., Ramalli, E., & Pernici, B. (2024). Towards a policy tuning method for data ecosystems. In *2024 IEEE International Conference on Web Services (ICWS), pages 55–64*. IEEE.
Sapsford, R., & Jupp, V. (1996). *Data collection and analysis*. Sage.
Sarker, M. N. I., Kamruzzaman, M. M., Huq, M. E., Zaman, R., Hossain, B., & Khurshid, S. (2021). Smart city governance through big data: Transformation towards sustainability. In *2021 international conference of women in Data science at Taif University, WiDSTaif 2021*.
Scheider, S., Lauf, F., Geller, S., Möller, F., & Otto, B. (2023). Exploring design elements of personal data markets. *Electronic Markets, 33*(1), 28.
Schmeling, J., Al Dakruni, S., & Mergel, I. (2025). Data collaboration in digital government research: A literature review and research agenda. *Government Information Quarterly, 42*(3), Article 102063.
Singh, D. (2020). Towards data privacy and security framework in big data governance. *International Journal of Software Engineering and Computer Systems, 6*(1), 41–51.
Stalla-Bourdillon, S., Wintour, A., & Carmichael, L. (2019). *Building Trust Through Data Foundations A Call for a Data Governance Model to Support Trustworthy Data Sharing. page 33*.
Susha, I., Flipsen, M., Agahari, W., & Reuve, d. (2020). *Towards Generic Business Models of Intermediaries in Data Collaboratives.pdf*.
Susha, I., Grönlund, Å., & Van Tulder, R. (2019). Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions. *Government Information Quarterly, 36*(1), 112–128.
Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. In *Proceedings of the 50th Hawaii international conference on system sciences (2017)*. Hawaii International Conference on System Sciences.
Susha, I., van den Broek, T., van Veenstra, A.-F., & Linåker, J. (2022). An ecosystem perspective on developing data collaboratives for addressing societal issues: The role of conveners. *Government Information Quarterly, 40*, Article 101763. data collaboratives sustainability.
The New Hanse Project. (2023). *Governing urban data for the public interest*.
Urbano, V. M., Bartolomucci, F., & Azzone, G. (2024). Determinants for university students' location data sharing with public institutions during COVID-19: The Italian case. *Data & Policy, 6*.
Verhulst, S., & Sangokoya, D. (2015). *Mapping the next frontier of open Data: Corporate Data sharing | by Stefaan G. Verhulst | internet monitor 2014: Data and privacy | medium*.
Walsham, G. (1995). Interpretive case studies in is research: nature and method. *European Journal of Information Systems, 4*(2), 74–81.
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., , … Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1), 1–9.
Williams, S. (2020). *Data action using data for public good, volume 3*.

Yang, T.-M., & Maxwell, T. A. (2011). Information-sharing in public organizations: A literature review of interpersonal, intra-organizational and inter-organizational success factors. *Government Information Quarterly, 28*(2), 164–175.

Yeh, Y. T., Eden, R., Fielt, E., & Syed, R. (2025). The role of use for the business value of big data analytics. *Journal of Strategic Information Systems, 34*(2), Article 101888.

Yin, R. K. (2018). *Case study research and applications, volume 6*. CA: Sage Thousand Oaks.

Federico Bartolomucci is Junior Assistant Professor at at the Department of Management, Economics and Industrial Engineering at Politecnico di Milano. He holds both a Bachelor's and a Master's degree in Management Engineering, and earned his Ph.D. in Management Engineering from Politecnico di Milano in 2024. His research focuses on data sharing configurations, collaborative innovation environments and social innovation. He has extensive experience in managing international research projects, and he has been co-chair of the EU Commission Special Group writing the Code of Conduct for the Use of Data in the Social Economy.

Edoardo Ramalli is a researcher at Joint Research Centre of the European Commission. He holds both a Bachelor's and a Master's degree in Computer science and engineering, and earned a Ph.D. in Information Technology from Politecnico di Milano in 2024. His research lies at the intersection of data ecosystems and artificial intelligence, with a strong focus on data management and knowledge extraction. He is particularly interested in the development of agentic systems and the integration of large language models to support intelligent, context-aware decision-making.

Valeria M. Urbano is a Junior Assistant Professor at the Department of Management, Economics and Industrial Engineering at Politecnico di Milano. She holds both a Bachelor's and a Master's degree in Management Engineering, and earned her Ph.D. in Data Analytics and Decision Sciences from Politecnico di Milano in 2023. Her research focuses on public transportation management, healthcare systems, and data sharing in the public sector, employing data-driven methodologies to support decision-making and service innovation. She has experience in inter-organizational governance in the healthcare sector, gained through participation in European projects.