Information and Organization 36 (2026) 100625

Contents lists available at ScienceDirect

Information and Organization

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

Stances toward data governance: Negotiating tensions in data sharing for artificial intelligence in healthcare

Rahel Gubser [a,c], Matthias Schulte-Althoff [a], Jörg Pohle [b], Nils Heinemann [b], Felix Balzer [a], Daniel Fürstenau [a,c,*]

[a] *Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Informatics, Berlin, Germany*
[b] *Alexander von Humboldt Institute for Internet and Society (HIIG), Französische Straße 9, 10117 Berlin, Germany*
[c] *Freie Universität Berlin, School of Business & Economics, Garystr. 21, 14195 Berlin, Germany*

A R T I C L E I N F O

Keywords:
Data governance stances
Health data sharing
Data governance practices
AI development

A B S T R A C T

Sharing data to enable artificial intelligence (AI) development creates unique tensions in contemporary data governance: balancing model robustness and bias reduction with strict requirements to protect privacy and personal rights, defining conditions for responsible data reuse, and reconciling control over data with desires for open data. In healthcare and nursing, these tensions are particularly pronounced, and existing governance arrangements often prove insufficient to resolve them. However, we still know little about how data-sharing initiatives actually navigate these tensions in practice, and how their governance activities form into overarching orientations, or stances, toward data governance. This study examines four forerunner initiatives that sought to establish data repositories to share data for future AI development. A comparative case analysis reveals twelve concrete data governance practices – situated activities through which stakeholders addressed the specific challenges of sharing data. Patterns across cases revealed that these practices formed three overarching stances toward data governance. We present a progressive, integrative, and protective stance toward data governance, each highlighting specific practices of reconciliation and leading to a variety of outcomes for future AI development. Our findings highlight the importance of aligning normative, organizational, and technical practices and suggest that data governance for sharing data to enable AI development is far from uniform. Instead, it encompasses diverse and situated practices through which responsibility, control, and openness are continually negotiated.

1. Introduction

The research field of data governance deals with balancing risks and the value of data use (Davidson et al., 2023; Vial, 2023). The need for such balancing can be understood as a response to deeply rooted tensions, in which legitimate interests in data use must be weighed against usage-related risks, for example regarding the protection of the privacy of individuals affected by data use in healthcare (Kearns & Roth, 2019; Sabatello et al., 2022; Seibert et al., 2021, 2023). Tensions describe enduring, interdependent demands that cannot be fully satisfied simultaneously (Putnam et al., 2016). In particular, such tensions become salient when data are to

---

* Corresponding author.
E-mail address: daniel.fuerstenau@fu-berlin.de (D. Fürstenau).

https://doi.org/10.1016/j.infoandorg.2026.100625
Received 6 August 2024; Received in revised form 24 April 2026; Accepted 28 April 2026
Available online 8 May 2026
1471-7727/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

be shared across organizational boundaries, where the rights of affected individuals cannot be sufficiently weighed against broad but not yet clearly defined future usage scenarios, as is often the case when sharing health data (Paparova et al., 2023; Winter & Davidson, 2019).

Considering the future data use for artificial intelligence (AI) development, a key issue is that developers often need to access data across multiple sites and require large, diverse sets of patient data to train robust and unbiased applications. Yet this need frequently clashes with the strict data protection regulations and privacy practices in healthcare, which are designed to safeguard patient confidentiality and uphold informed consent. As a result, conventional data governance practices can come into tension with AI developers' requirements for open and extensive data access. While the AI community emphasizes the benefits of broad data access as a prerequisite for progress, research shows that potential data contributors weigh these collective gains against individual risks such as re-identification, reputational harm, and unclear ownership or liability (Pujol Priego et al., 2022).

In the area of data governance, different research streams emphasize distinct aspects of balancing risks and the value of data use, including institutional and organizational perspectives focusing on governance structures and coordination (Davidson et al., 2023; Winter & Davidson, 2019), value-centric perspectives highlighting value creation and innovation (Kilgus et al., 2024; Lefebvre et al., 2025), and semiotic perspectives emphasizing the contextual meaning and evolving value of data (Benfeldt & Persson, 2025; Jarvenpaa & Essén, 2023). Extant research highlights the need to balance value creation and risk mitigation but provides limited understanding of how tensions are continuously negotiated across organizational boundaries, particularly in AI-oriented data sharing contexts. In this way, data governance primarily oriented toward risk mitigation may itself become part of the problem when it restricts value-generating data reuse, particularly when future use scenarios remain difficult to anticipate.

The purpose of this paper is therefore to *examine how data-sharing initiatives enact data governance to reconcile tensions in sharing data for future AI development.*

We take a situated perspective on data governance. We define data governance as a multidimensional reconciliation process through which stakeholders negotiate how data are collected, processed, shared, and reused, while distributing both value and risk (von Grafenstein, 2022). Data governance practices constitute the means through which this reconciliation is enacted, encompassing normative, organizational, and technological arrangements that shape how this process is carried out. To understand how this reconciliation process unfolds over time, we focus on the emergent, dynamic activities through which data governance is negotiated and through which patterns of practices give rise to what we conceptualize as stances toward data governance. A stance in this context is defined as an emergent overarching orientation that influences how stakeholders prioritize and implement governance practices to address tensions in data sharing for AI development in healthcare (cf., Fayard et al., 2016).

Our research context is Germany, where a federal funding program for research and development of data repositories and AI-based solutions for nursing care was launched in 2021. Using a qualitative case study approach (Gerring, 2007), we identified and analyzed practices of data governance within four similarly funded AI in nursing care data-sharing initiatives. We observed that they adopted three distinct stances toward data governance. Our results show that governance tensions are not resolved by single rules or structures but through situated negotiation and the combination of complementary practices.

Our contribution is: First, we advance current understandings of data governance by showing that organizations differ in how they approach data governance when sharing data for AI development. In doing so, second, we highlight twelve distinct data governance practices that address key tensions of data sharing for AI development, within areas of tension related to data scope, conditions of data use, and data access. This supports, third, the identification of three distinct stances toward data governance for data sharing: Progressive, Integrative, and Protective. By comparing initiatives across these stances, we highlight practical ways to address governance challenges and propose options for sustaining data infrastructures relevant for practitioners and policymakers in the context of AI development. Attention to such challenges informs the practical, yet fundamentally legitimate, use and establishment of shared data in healthcare beyond the current insistence of the European Union and national funding bodies on technical repositories and the predominant practical reality of data silos.

2. Theoretical background

This section engages with theoretical debates on the governance of data, particularly when the data needs to be governed because it is shared to support future AI development. We explain how we use a situated perspective that enables us to see data governance as patterned bundles of practices that adjust flexibly over time (Jarvenpaa & Essén, 2023) and across contexts (Benfeldt & Persson, 2025; Foster et al., 2018), and how this gives rise to stances that guide how actors prioritize and implement practices in addressing tensions in data sharing for AI development.

2.1. Governing the sharing of data to enable future AI development

The use of sensitive patient data for future AI development inherently requires thorough governance. This governance includes the governance of information technology systems used within such endeavors (Weill & Ross, 2004, 2005) as well as the governance of data that is needed for corpus collection and pre-processing, pre-training, training, fine-tuning, model evaluation, and similar activities.

The governance of data increasingly extends *beyond* intra-organizational boundaries and hierarchies of control (Davidson et al., 2023; Khatri & Brown, 2010; Paparova et al., 2023) to collaborative settings where actors need to work together (Winter et al., 2014). Here, data governance evolves into a negotiation process (Jussen et al., 2024; Lis & Otto, 2020). For example, data-sharing initiatives may need to provide evidence and arguments for multiple hospital decision makers, such as data protection officers, ethics committees, and other actors before patient data can be shared across several sites.

This negotiation process considers tensions in reconciling value creation and potential risks of sharing data. In data-sharing initiatives, stakeholder risks are recognized in prior literature: Data providers may fear a competitive disadvantage (Klievink et al., 2018; Zeiringer & Thalmann, 2022), a lack of credit, a lack of rewards, misuse, misinterpretation, or liability concerns (Pujol Priego et al., 2022), while data subjects – patients – or society are concerned about privacy breaches when (personal) health information is inappropriately repurposed in new contexts (Winter & Davidson, 2019). While these risks are grounded in legal rights, the potential value of sharing data for future AI development can remain abstract for the individual or organization who is asked to share (von Grafenstein, 2022). For example, it may take years before a new AI algorithm is developed, until it is packaged into a product, and it may not benefit the same patient (group) that originally provided its data. Because the benefits of sharing often materialize only later and may accrue to actors other than those who bear the immediate liability risks (Pujol Priego et al., 2022), conventional data governance models are often inadequate for AI-oriented data sharing (von Grafenstein, 2022). In such settings, no single organization can simply impose the rules for data use. Governance instead depends on inter-organizational negotiation over how value and risk should be distributed (Jussen et al., 2024).

Tensions in data governance when sharing data have been described from at least three perspectives (Table 1). The *organizational and institutional perspective* puts emphasis on institutions, structural forms, practices, and processes (e.g., Davidson et al., 2023; Winter & Davidson, 2019). According to this view, tensions in data sharing may come from unbalanced business priorities vis-à-vis personal rights and societal and public interests (Susha et al., 2023; Winter & Davidson, 2019, 2022). Initiatives need to govern data to balance benefits and risks and need to cope with several hurdles, such as missing incentives (Susha et al., 2023), missing infrastructures (Susha et al., 2023), and insufficient participation structures (e.g., Winter & Davidson, 2019). Organizations may cope by setting up appropriate organizational, regulatory, and technical/ algorithmic data governance forms for the sharing of data (Winter & Davidson, 2019; see Appendix 1). This literature has also discussed how data sharing can break down when institutions to govern data are insufficient and individuals or the public lose trust in a data sharing arrangement (Winter, 2021; Winter & Davidson, 2019).

While this stream, by describing different forms of data governance (e.g., Paparova et al., 2023; Susha et al., 2023; Winter & Davidson, 2019), has provided important groundwork for a more holistic understanding of data governance and thereby points to the limitations of a purely organizational perspective, the explanatory aspect of how institutionalized forms of data governance build or erode trust in a particular data sharing arrangement remains insufficiently developed. Moreover, data governance as a continuous process of negotiation is suggested but not fully and comparatively demonstrated, which provides a point of departure for our study.

Table 1
Perspectives on data governance and data sharing.

Literature Stream	Focus	Core Tensions	Mitigations	Key References
Organizational & Institutional Perspective	Focuses on formal structures, institutions, roles, and processes governing data sharing (e.g., policies, compliance, accountability mechanisms).	- Balancing organizational/ business value vs. individual rights and public interest - Lack of trust in data sharing arrangements - Risk of breakdown when governance is perceived as insufficient	- Establishing organizational, regulatory, and technical governance frameworks - Creating clear roles, responsibilities, and accountability structures - Implementing compliance mechanisms (e.g., ethics boards, data protection officers) - Building institutional trust	Davidson et al. (2023); Winter and Davidson (2019); Paparova et al. (2023); Susha et al. (2023)
Value-Centric / Economic Perspective	Emphasizes value creation, incentives, and scalability of data sharing; treats data as an economic asset.	- Trade-offs between data quality, accessibility, and privacy - Uneven distribution of costs, risks, and benefits - Underestimation of legitimacy and ethical concerns - Tensions when technical feasibility exists but social acceptance does not.	- Designing technical infrastructures (e.g., federated learning, decentralized architectures) - Implementing data access controls and governance mechanisms - Creating incentive structures (e.g., shared ownership, governance bodies) - Ensuring data quality management and standardization	Dyke et al. (2018); Jarke and Quix (2022); Lefebvre et al. (2025) Kari et al. (2025) Kilgus et al. (2024) Parsons et al. (2025); Tiwana et al. (2010) Wixom et al. (2020)
Semiotic Mediation Perspective	Focuses on meaning-making processes and how data acquires value across contexts and over time (iconic, indexical, symbolic dimensions).	- Shifting meanings of data across contexts - Risks from rigid or overly flexible interpretations - Tension between interpretability and reuse risks (e.g., re-identification, misinterpretation) - Difficulty translating meaning into actionable governance decisions	- Developing context-sensitive interpretation practices - Maintaining flexibility in data representation and reuse - Linking meaning-making to governance mechanisms (e.g., access decisions, scope definitions) - Supporting interpretive alignment across stakeholders	Alaimo and Kallinikos (2024) Benfeldt and Persson (2025) Aaltonen et al. (2021) Aaltonen et al. (2023) Jarvenpaa and Essén (2023)

The *value-centric, economic perspective* is business outcome-driven and has taken a more incentive-oriented stance. It often starts by assuming that organizations need to share data because AI development will need high-quality, well-structured data. This makes data repositories, data spaces, and similar arrangements, from diverse sources such as real-world data, essential for enabling AI development (e.g., Friedrich et al., 2026; Kilgus et al., 2024; Lefebvre et al., 2025; Wixom et al., 2020). This literature has acknowledged that a missing focus on data quality or missing data governance could undermine the value of data and has suggested that data sharing arrangement need to balance data quality, accessibility, and privacy to create business value (Jussen et al., 2024; Lukyanenko et al., 2019; Parsons et al., 2025). According to this view, new technical arrangements may be needed to enable the responsible repurposing of data, such as federated or decentralized data integration (Jarke & Quix, 2022; Prinz et al., 2022; Tiwana et al., 2010), data usage controls (Jung & Dörr, 2022), or controlled access models (Dyke et al., 2018). Further, organizational arrangements may also be needed, such as the involvement of a committee to release the data, as well as new organizational governance forms considering shared ownership of newly emerging organizational entities (Kari et al., 2025).

This stream offers an important shift in perspective, as it places less emphasis on risks and instead understands value creation as an equally central component of data governance. As a result, incentives come more strongly into focus, providing concrete entry points for more balanced data governance arrangements. However, this perspective remains insufficiently developed in accounting for how deeply entrenched some of the tensions in data governance can be, how "fragile" certain achieved compromises may be, and how crucial ongoing negotiation is for the success of specific data-sharing arrangements. Moreover, economic interests tend to be comparatively overemphasized, while other social interests are treated as subordinate, which we seek to bring more centrally into the analysis.

Finally, the *semiotic mediation perspective* highlights the meaning-making process through which stakeholders can make data into a valuable asset across contexts (Benfeldt & Persson, 2025; Foster et al., 2018) and over time (Jarvenpaa & Essén, 2023). This semiotic perspective includes iconic, indexical, and symbolic dimensions of signification. The iconic dimension refers to representations that resemble real-world phenomena, such as a heart symbol in a health app. Yet this resemblance can also backfire when the coupling between sign and reality becomes too rigid – such as when a simple female symbol no longer adequately represents gender diversity (Aaltonen et al., 2021.; Benfeldt & Persson, 2025). The indexical dimension concerns direct referential links between data and the real world, for example, when a birthdate is encoded in a national identification number. While such mappings enhance readability and efficiency, they can reduce interpretative flexibility and amplify risks when data are reused across numerous systems (Benfeldt & Persson, 2025). Lastly, the symbolic dimension reflects how shared norms, cultures, or conventions shape interpretation – such as the convention that "green" in a traffic light means "go."

The semiotic stream is valuable because it shows that data do not travel with fixed meaning; however, it requires more explanation of how stakeholders translate shifting meanings into concrete governance decisions about scope and the legal bases for reuse, including access rights (Paparova et al., 2023). In AI-oriented sharing, initiatives need to connect meaning-making to operational choices (Papagiannidis et al., 2025). Put differently, meaning must be converted into governable action considering the contextual and social processes through which the data are interpreted. This motivates a situated perspective that connects meaning with legitimacy and technical constraints (Davidson et al., 2023).

In conclusion, the literature on data governance provides us with important insights into central tensions of governing data sharing and respective mitigations. However, especially when the context is data sharing for future AI development in healthcare, the literature remains inadequate to fully capture the range of issues present. The organizational and institutional stream already emphasizes the institutional conditions that give rise to trust in data-sharing arrangements (including for AI development), but it pays insufficient attention to the orientations of actors that, from a practical perspective, normatively guide action and, through repeated negotiations, may lead to the erosion or reinforcement of such trust. The economic stream foregrounds scale and incentives but is less explicit about how initiatives secure legitimacy and negotiate tensions when data for AI development may be technically valuable yet normatively contested because risks, benefits, and future uses remain unevenly distributed (Kilgus et al., 2024; Lefebvre et al., 2025; Parsons et al., 2025). The semiotic stream shows how the meaning and value of data shift across contexts and over time, but offers less guidance on how actors translate these shifting interpretations into concrete governance decisions about scope, reuse, and access for AI development (Benfeldt & Persson, 2025; Foster et al., 2018; Jarvenpaa & Essén, 2023). A situated perspective helps address these limitations by focusing on governance as an ongoing accomplishment through which actors negotiate, adapt, and combine practices to reconcile tensions in data sharing for future AI development (Abraham et al., 2019; Benfeldt Nielsen, 2017; Jussen et al., 2024; von Grafenstein, 2022).

2.2. Advancing a situated perspective on data governance

2.2.1. Data sharing tensions

Our study takes as its starting point the conceptualization of central tensions that arise in collaborative data-sharing settings for future AI development. Building on Putnam et al. (2016), we conceptualize tensions as trade-offs between interdependent stakeholder demands. We investigate and conceptualize these demands across three main areas (von Grafenstein, 2022): tensions arising around (i) *which and how many data can be used for sharing and use in AI development (Data Scope)*, (ii) *how restrictive the conditions are under which this data can be used and reused (Conditions of Use)*, and (iii) *who can access the data and for what purposes (Data Access)*. These distinct yet interrelated areas of tension appeared repetitively in the empirical material. At the same time, they follow the practical sequence through which data-sharing arrangements are negotiated and enacted across the data pipeline. The areas of tension serve as key reference points for governance and illustrate the areas of conflict that emerge when balancing risks and values in sharing data for AI development.

Data Scope concerns the issue of making available large, representative data sets as well as reducing the data to protect data subjects from being exposed to risks from re-identification or misuse. *Conditions of Use* revolves around the permissible uses of the data beyond the purpose for which it was originally collected – conditions range from narrow, use-specific to wide, use-agnostic interpretations (Parsons et al., 2025). *Data Access* concerns whether data are made openly available or tightly controlled, reflecting the push-pull between openness in favor of AI development on the one hand and control to protect the autonomy and competitive interests of data providers on the other.

2.2.2. Data governance practices

We refer to *practices* of stakeholders as the things that they say or do in a routine way and emphasize the often taken-for-granted nature of our social activities. By advancing a situated perspective, we respond to the lack of practice-based methods in data governance research that has been noted as a critical limitation, calling for the attention to ground-level of data governance (Benfeldt Nielsen, 2017; Benfeldt & Persson, 2025; Lefebvre et al., 2025). These practices may differ in settings that are highly institutionalized such as healthcare, where certain values are held professionally, normatively, or economically, such as the sensitive nature of the data and implications of potential misuses. Successful data governance practices are defined by the extent to which they foster alignment and reconcile conflicting interests in a productive and resolutive manner (Foster et al., 2018).

Our situated perspective also emphasizes the *iterative*, *processual*, and *tentative* nature of data governance (Benfeldt Nielsen, 2017). This involves turning attention to the situated actions, negotiations, and adaptation processes that have been emphasized before (e.g., Abraham et al., 2019; Jarvenpaa & Essén, 2023; Jussen et al., 2024) but that were not very often examined.

Our interest in practices extends the current literature around *organizational, normative*, and *technical* practices of data governance that create observable responses to emerging tensions in sharing data for future AI development. The core of *organizational practices* lies in using structural, relational, and procedural arrangements (such as roles, committees, and processes) to coordinate and control how data are governed. The core of *normative practices* lies in translating regulatory requirements and moral principles into expectations and obligations for actors. The core of *technical practices* lies in embedding rules, constraints, or permissions into data, algorithms, and code, shaping what actors can or cannot do with data. Table A1 in the appendix shows exemplary practices from prior literature that address traditional data governance issues in the areas of scope, conditions of data use, and access.

In turn, we do not study data governance as something that is "designed" in a boardroom (e.g., Black et al., 2023) or that consists of a single rule or policy, but as an ongoing accomplishment of the involved stakeholders on the ground.

Our interest in data governance as situated action has originally evolved from von Grafenstein's (2022) work as a legal scholar, who had advanced a definition of data governance as a multidimensional reconciliation process. In essence, reconciliation means that arising tensions – such as which and how many data to use, when to (re)use them, and who can access data – will create the need for communication and often alignment between stakeholders.

From this vantage point, we suggest defining data governance as *practices of reconciliation among multiple stakeholders concerning the collection, use, sharing, and reuse of data*. Thus, data governance is understood as a process that aims to *reconcile conflicting interests in data among stakeholders related to the creation and distribution of both value and risks* (von Grafenstein, 2022).

2.2.3. Stances toward data governance

Finally, it appeared useful to us to highlight emerging overall orientations toward data governance. Borrowing the term stance in the sense of a collectively enacted orientation (cf. Fayard et al., 2016), we define stance to denote an emergent yet relatively stable inter-organizational orientation that is enacted through a relatively consistent configuration of practices. In this sense, a stance is broader than a single practice, more durable than a situational response, and more enacted than a formal policy statement.

Our main theoretical anchor is the literature on network governance (Sydow & Windeler, 1998), which draws on structurationist arguments to explain how practices become stabilized into overarching orientations that, in turn, guide action. This action-guiding capacity has multiple sources; in particular, normative, resource-based, and signifying guiding principles are emphasized theoretically, all of which also play an important role in our case.

In doing so, we see data governance, similar to Winter and Davidson (2019), as interlinked sets of more dynamic, role-sensitive reconciliation processes that make governance actionable and result in sustainable outcomes that favor AI development over time (Jarvenpaa & Essén, 2023). For example, when data-sharing initiatives lean toward open sharing, they may need to investigate and combine practices to prevent re-identification in the present and over time because traditional anonymization methods may not be sufficient. In doing so, a stance represents an overall orientation toward data governance. Stances emerge from the combination of normative, organizational, and technical practices, yet they primarily express a relatively stable collective attitude toward governance that exceeds the sum of these activities. This attitude evolves through, and in turn shapes, the negotiation of governance processes. Table 2 summarizes our conceptualization.

3. Methods

A multiple case study was conducted (Gerring, 2007) to examine how data-sharing initiatives enact data governance. The research aims to develop theory (i.e., in the sense of Sandberg & Alvesson, 2020) on the categorization and interdependencies of data governance practices to address *multiple tensions* in governing data sharing for AI development. We seek to provide insights from data-sharing initiatives by considering data governance practices as focal points of conceiving and enacting (*data governance stances*), offering a better understanding of how conflicting stakeholder interests are reconciled.

Table 2
Core concepts of our situated perspective on data governance practices.

Core Concept	Definition	Key references
Data Sharing Tensions	Interdependent demands of stakeholders that involve trade-offs (Putnam et al., 2016, p. 73) considering a) which and how many data to share/use, b) under which conditions to share/(re)use, and c) who can access and use and for what purposes.	Putnam et al., 2016 von Grafenstein, 2022
Data Governance Practices	What stakeholders say or do to address prevalent tensions. Practices are placed in context and studied in time. They can be distinguished in organizational, normative, or technical-algorithmic dimensions of data governance.	Barley, 1986 Winter and Davidson, 2019 von Grafenstein, 2022
Stances Toward Data Governance	The emergent inter-organizational orientations toward data governance, which influence how stakeholders prioritize and implement practices that make data governance actionable and result in sustainable outcomes for AI development.	Fayard et al., 2016 Sydow & Windeler, 1998 Davidson & Winter, 2019

3.1. Research process and setting

Four cases – Alpha, Beta, Gamma, and Delta – were selected. The cases are multi-stakeholder research projects aiming to make patient data accessible to researchers with the goal of fostering AI development and building the foundation for future AI solutions in nursing. The leading institutions driving the cases were university clinics, medical device or system vendor companies, and care institutions located in Germany. The sources of data for processing and sharing mostly stemmed from electronic medical and nursing records (Alpha, Beta, Gamma), and in one case (Delta), involved sensor data (Table 3).

Following Gerring (2007), the cases took on different roles. Beta and Gamma were considered *typical* cases as they followed a public infrastructure model and were strongly connected to established, publicly funded data-sharing infrastructures. Alpha and Delta were more *deviant*: Alpha because it took a stronger technology-driven approach, yet, within a public institutional setting, driven by a university clinic and a machine learning institute; Delta because it was driven by a set of private companies with strong market penetration ambitions.

The cases offered a good basis for comparison because they were all selected in a competitive selection procedure for public funding to build data repositories for AI in nursing, involving significant legal and political issues and impactful stakeholders (average funding sum per project: 2.31 million Euros). As such, these cases were diverse in their pursued approaches to building data-sharing infrastructure but shared the funding framework under which they operated, promising rich and nuanced empirical insights. Operating under German and EU laws, they allowed a comparative analysis, marked by a pre-selection process, limited project durations (24–36 months), and financial constraints.

Our access to the four chosen cases was facilitated through our involvement in an overarching project associated with the entire funding line. Notably, our deep engagement with Alpha as part of the project team provided us with special access to interviewees, documents, and materials, enriching our analysis.

3.2. Data collection

We collected data from multiple sources (see Appendix 2), spanning a timeframe from the beginning of the funding line in November 2021 until April 2025. This approach helped us collect detailed information, including perspectives and general views on data governance. It also helped us understand the context and the cases.

Subsequently, for each case, we began collecting data through semi-structured expert interviews (Meuser & Nagel, 2009). We engaged with several key stakeholders (e.g., industry partners, project coordinators, data governance specialists, nursing scientists, technical partners), exploring the data governance of the data-sharing initiative for underlying causes and actions following decisions. Additionally, we interviewed “data-sharing enablers” such as governance experts, data infrastructure builders, economists, and representatives of a German data-sharing consortium we refer to as ‘parent organization’ (Appendix 3 offers an overview). We conducted 44 interviews,[1] with an average length of 39 min, from October 2022 to April 2025, following guidelines for semi-structured interviews (Spradley, 1979), and innumerable informal exchanges throughout the entire project duration. Appendix 4 lists the interview guides. All interviews were recorded and transcribed, and we also took detailed notes for later discussion. This process was iterative, with the findings refined after each round of coding and discussion.

In addition, for our in-depth analysis, we collected archival data from each emerging initiative using 202 pages of internal project plans and 134 pages of governance guidelines from parent organizations for cross-validation purposes. Documents also included data protection plans, consent forms, internal presentations, and communication memos.

Our understanding of the cases was informed by careful reading of the material (Mayring & Fenzl, 2014). Furthermore, we participated in several seminars on data protection attended by representatives from all cases, followed by an on-site discussion of legal requirements and decision-making. Finally, in-depth observations for one case (Alpha) were possible, including access to all meetings

---

[1] 8 for case #1, 5 for case #2, 11 for case #3, 6 for case #4, 8 for parent organizations, and 6 for other experts on data sharing

Table 3
Overview of cases selected for multiple case study.

Cases	Case 1 Alpha	Case 2 Beta	Case 3 Gamma	Case 4 Delta
Goal[1]	Making available nursing data in data repository to support AI development			
AI Development	Develop a clinical decision support for nurses using Machine Learning prediction models	Develop clinical decision support in intersectoral settings using Machine Learning prediction models	Develop clinical decision support in ICU for nurses and doctors using Machine Learning prediction models	Establish a community of data users that get access to secretion data for AI-based research
Key Use Case	Fall prevention	Fall prevention	Delirium prediction and management	Secretion management
Data	Electronic nursing data (medical diagnoses, procedures, medication	Electronic nursing data and manual assessment data from care homes	Electronic nursing data (intensive care / vital parameters) and manual nursing documentation	Vital sign sensor data from specific medical equipment (i.e., ventilator machines); associated meta data.
Project Period	Aug 22 – Aug 25	March 22—Feb 25	March 22—Jan 25	March 22—Feb 25
Key Stakeholders				
Data Subjects	Retrospective patients	Prospective + retrospective, patients	Retrospective, patients	Prospective + retrospective, patients
Data Provider	University Clinic 1, Geriatric Clinic 2	Care home 1	University Clinic 2	Care home 2, medical device company
Data Processor	The infrastructure provider (project)	The infrastructure provider (project)	Data is hosted on an infrastructure provided by parent organization	Data is hosted by medical device company
Data Recipient	Internal nurses, external researchers + ML developers	Internal nurses, external researchers + ML developers	Internal researchers primarily	External ML developers + researchers

[1] The goal of each initiative was predetermined by the funding line, allowing for a high degree of comparability across cases.

(weekly meetings, workshops, technical discussions, etc.), notes, presentations, and further documents (e.g., study plans, work package progress reports, weekly meeting notes, data protection plans).

3.3. Data analysis

Our analysis unfolded in five iterative cycles, each deepening the link between empirical material and theory (i.e., Rouse et al., 2025).

3.3.1. Case exploration

We began with a close, open-ended engagement with three cases. Guided by von Grafenstein's (2022) multidimensional view of data governance and by a general interest in the cases' overarching vision, their anticipated technical solution and economic exploitation ideas, we exploratorily coded initial interviews, project plans, and meeting notes. Regulatory documents and stakeholder maps informed our case understandings. This round surfaced three recurring ways of organizing data sharing, which we provisionally labelled ‘data governance stances’. Following Sandberg and Alvesson's (2020) typologizing logic, these early categories served as sensitizing concepts rather than final constructs.

3.3.2. Expanded data collection and practice coding

After adding the fourth case and additional interviews, we focused on the concrete “doings and sayings” (Schatzki, 2002). Drawing on Barley's (1986) notion of technology–enabled structuring, we identified 18 distinct data governance practices and compiled key quotations for each. In particular, we noted the difference between simply “working with data” (e.g., Parmiggiani et al., 2022) and making decisions about data-sharing benefits versus risks. The latter involves creating responsibility schemes around data (a data governance practice). This distinction was vital in guiding our analysis. An example of a code is “responsibility stays with data-sharing initiatives” (with empirical excerpts including: “I strongly believe that we should take responsibility for our own actions. If we decide that data is relevant, we must justify this decision well, make the potential risks transparent, and explain how we will deal with them”). MAXQDA supported both open and selective coding, while the analysis of practices remained anchored in the normative, organizational, and technical categories highlighted by von Grafenstein (2022). This step also involved cross-case validation of identified practices across cases. The findings from the initial case, Alpha, served as a baseline for categorizing the subsequent cases. We assessed the recurrence and relevance of practices across cases by systematically comparing their presence and operationalization. Central questions included: Is this practice observable in other cases? Are there additional practices not initially evident in Alpha?

3.3.3. Practice consolidation

Using Grodal et al.'s (2021) principle of parsimony, we combined overlapping codes. Focusing on parsimony reduced our list from 18 to 12 practices while keeping empirical richness (Appendix 5). We also recognized a difference between implemented (explicitly present) and conceptualized practices. This clarity helped us to create a categorization scheme. We distinguished between: *a*) practices that were implemented, meaning they were present in the data and the case, b) practices that were discussed but not implemented, and c) practices absent from a case. The distinction between conceptualized and implemented practices guided our next step, considering interrelations between practices. ***Over time,*** it became evident what practices were practically relevant to address tensions, while others remained conceptual. This allowed us to deepen our understanding of the temporality of data governance and prepared our next step to converge practices into stances.

3.3.4. Converging practices into stances

We mapped every practice to the emerging stances, refining labels for conceptual clarity (see Table 5). We first identified the essential practices required to establish each initial stance and left only the practices necessary to enact it. For instance, sharing data without consent under a *research privilege* requires additional measures: the integrative stance retains data locally, whereas the progressive stance transforms data on a technical dimension further to prevent future re-identification. Two cases that displayed near-identical practice configurations were combined under one approach. This process yielded three data governance stances: progressive, integrative, and protective, each conceptualized as emergent configurations whose internal practices and cross-case patterns were compared (see Appendix 6). Finally, we examined how each stance materialized, either as distinct forms or along a spectrum (e.g., from uniform to heterogeneous data).

3.3.5. Reconceptualizing core tensions

A targeted literature review on data governance in data sharing for AI development and team discussions during regular meetings clarified three foundational tensions – scope of data, conditions of use, and access rights – around which all practices were clustered. This step sharpened the explanatory power of the situated perspective. Throughout the process, the authors met regularly to challenge code definitions, compare memos and resolve discrepancies, ensuring analytical rigor and shared interpretation. We also reached out in February and July 2025 and obtained feedback from all four cases on the obtained data governance practices, ensuring that our observations were in line with participants' perceptions.

3.4. The cases

All cases represent initiatives to share data in nursing to enable future AI development. The cases consist of various comparable stakeholders from public and private institutions with access to patient data. For a case overview, see Table 3.

3.4.1. Alpha

This project was a collaboration between two primary institutions: an urban university hospital (Clinic 1) and a specialized geriatric clinic (Clinic 2). Clinic 1, with its extensive infrastructure for research and clinical trials, provided data from 50 wards, while Clinic 2 contributed documentation from geriatric stationary stays, where fall prevention is a central aspect of daily care. The project aimed to build an open yet secure data infrastructure that would allow researchers and innovators – such as developers of machine learning models – to access relevant clinical data. In practice, this raised several challenges. For example, Clinic 2 had only CSV data dumps of two versions of the electronic documentation system, whereas Clinic 1 worked with sophisticated electronic health records but had to connect several nursing–relevant tables to their health data platform to export the data while ensuring that legally protected information was not revealed. The combined data set contained routine care documentation in tabular and textual form, including diagnoses, nursing notes, medical plans, and mobility and cognitive assessments, used to train fall prediction models. Alpha's approach encouraged both clinics to share their data without physically transferring them. Instead, the project piloted a federated learning setup and data–synthetization methods, allowing model training on local servers. This setup aimed at demonstrating how privacy-preserving technologies can enable AI development and open data sharing even in sensitive hospital environments, while simultaneously highlighting the practical difficulties of aligning data formats, governance procedures, and ethical and legal requirements across institutions.

3.4.2. Beta

The Beta project was led by a research institute specializing in data and software systems, focusing on interoperable information infrastructure and data sharing architectures, and included a large provider of electronic care documentation systems and a long-term care provider. The project addressed the challenge of standardizing and thus allowing for the utilization of heterogeneous patient data, specifically structured data from ambulatory, day care, and home care. The project's initiative to harmonize this data aimed to mitigate the issue of disparate data formats, e.g., when a cognitively impaired person is transferred from a hospital to home care, that hinders the integration and repurposing of data for effective data analysis and evaluation. By creating a centralized data repository, the Beta project aimed to facilitate secure and legitimate access to nursing care data for further processing and interpretation through connecting this data to Germany's medical informatics initiative data access platform (parent organization). Beta focused on the same use case (fall prevention), facilitating straightforward comparison between cases. The approach used participatory formats, including workshops and co-design sessions with nursing staff. Beta's data sharing concept considered the use of a structured, standardized data format.

3.4.3. Gamma

The third initiative, Gamma, anticipated developing a hybrid AI-supported system for identifying delirium risk, providing practical decision-making support for nursing professionals to address the challenge of this acute but reversible cognitive impairment. The system's underlying data foundation integrated various data sources, enabling pattern recognition for recommendations. Like Alpha, this initiative was affiliated with a German University Clinic 3. However, the two cases differ significantly in terms of their data governance. Just like Beta, Gamma foresaw sharing data with an established consortium, Germany's Medical Informatics Initiative. Furthermore, the initiative aimed to install a separate server on the hospital's premises, where a validated algorithm would also be integrated. Therefore, Gamma's data sharing concept considered a standardized, anonymized dataset, while model sharing enabled the use of AI systems within the clinic's separate infrastructure. Unlike Alpha and Beta, where the data focused on socio-demographic information, medical information (diagnoses and procedures), as well as nursing information (fall assessments), Gamma included additional monitoring data such as vital parameters from critical care information systems.

3.4.4. Delta

The Delta case brought together a public research center, medium-sized medical technology and IT firms, and a university clinic to tackle the challenge of siloed data in care homes, which included information from medical technology, sensors, and documentation. Delta aimed to create an independent infrastructure that integrates diverse data sources. Delta's plan was to onboard patients in care homes to donate their monitoring data. The final case stood out from the others in two main aspects: First, it involved a deeper partnership with private sector entities, especially medical device manufacturers. Second, it centered on data that included vital parameters indicative of sensitive care information yet stripped of any obvious personal identifiers. This data was used to optimize secretion management. Delta considered several steps in their data sharing plan: First, data from disparate systems would be aggregated as unprocessed data on a central platform; for this, patients had to provide informed consent. Data should then be automatically anonymized and processed on the platform, to incentivize competitive, private data providers to also share their data on the platform. More sensitive data such as personal information would remain on site, where federated learning would be applied and synthesized copies of such data would – in a last step – also be integrated into the platform for further use.

4. Results

Across the four initiatives, we identified three stances toward data governance – progressive, integrative, and protective – which reflect how actors negotiated tensions around data scope, conditions of use, and access. Each stance combines normative, organizational, and technical practices in different ways. The aggregated practices that the initiatives relied on are shown in Table 4. The following section will highlight some important results. Appendix 5 provides a detailed coding system.

4.1. Progressive stance: alpha negotiating the boundaries of openness

4.1.1. Data scope. At Alpha, the project team aimed to share a clinical patient dataset with an ambition to provide large-scale datasets for AI development. Alpha took advice in that regard from a scientific advisor: *"Research needs generalizable statements, [it] has to stratify, for example, by gender but also many other characteristics, and for that, it needs large amounts of data" (Scientific Advisor and Expert in Medical Informatics,* PO-A).

To train predictive models for fall risk, the team began by extracting data from hundreds of thousands of patient records, including medical, nursing, and medication variables. Using feature–importance techniques, the team examined which data fields most strongly influenced model performance. The analysis showed that fall history and core nursing assessments were reliable predictors, even when diagnostic detail was reduced. This finding directly informed later decisions about which data fields would be included in the trained models as well as in the planned open data repository (OD-24; P: *data scoping*). Alpha thereby managed to reduce the variety in their initial dataset, addressing regulatory data-minimization requirements without compromising model performance.

Over time, the project's data scope expanded as new researchers joined the team and new questions emerged, such as examining not only fall occurrence but also fall severity. Each addition required renewed negotiation between hospitals and data scientists over what could be shared. Eventually, the team of data scientists added 28 separate datasets within the internal data pipeline, each tailored to a specific use case and circle of recipients.[2]

To further expand data scope, Alpha kept partners engaged by introducing show-and-tell sessions where data analysts demonstrated early findings to clinicians from participating hospitals. When Clinic 1 shared preliminary analyses on nursing quality, Clinic 2 responded by offering additional data, hoping to use similar insights for staffing and resource planning. Such reciprocal exchanges illustrate how creating value for key stakeholders became a deliberate normative data governance practice that tied knowledge creation to data contribution (P: *luring with value*).

The materialization of luring with value was revealed at Clinic 2 with an intended shift in organizational culture: "*Over the past few years, hospitals have always focused very strongly on the individual patient... The focus is on the person and every patient is individual and we wanted to get out of this perspective and say: perhaps there are larger patterns that we can recognize if we process the data better.*" – CEO, Geriatrics Clinic II (Alpha-E). In response to the project, work profiles of nurses in Clinic 2 were revised. New data analysis tasks became integrated into the job profile, which presented a significant change to current practice. The process also raised data quality

---

[2] For example, one dataset contained complete medication records, including timestamps, for research on medication interactions and fall prediction – which increases the risk of re-identification. This dataset was made available to a data scientist at the same research institute, while the team decided that guest interns and students were denied access because medication data were not primary targets for their research.

Table 4
Data governance practices across areas of tension.

	1: Data Scope (Minimized - Maximized)				2: Conditions of Use (Narrow - Wide)					3: Data Access (Closed - Open)		
	DS	LV	EF	CQ	SD	RP	LP	BC	SS	OK	EC	LC
Alpha	●	●	●	○	●	●	○	●	●	●	☐	○
Beta	●	○	☐	☐	☐	●	●	○	●	☐	●	●
Gamma	☐	☐	○	●	☐	○	●	●	●	○	●	●
Delta	○	●	○	●	○	☐	☐	☐	○	○	○	○
Legend:	Data Scoping (DS) Luring with Value (LV) Evaluating Algorithmic Fairness (EF) Contextualizing Data Quality (CQ)				Synthesizing Data (SD) Exercising Research Privilege (RP) Broadening Consent (BC) Separating Storage (SS) Leveraging Parents (LP)					Opening Knowledge (OK) Engaging Communities (EC) Layering Access Control (LC)		

☐ – not present in the case; ○ – conceptualized, but not implemented in case; ● – implemented practice in the case.

awareness at Clinic 1 and Clinic 2, resulting in documented lessons learned for future projects and quality management (Workshop, February 2023).

Lastly, the team at Alpha evaluated fairness as part of its data governance. To balance model robustness with data-minimization requirements, Alpha aimed to anonymize and reduce the dataset intended for open sharing (OD-25). Alpha defined fairness as a measurable performance metric: They compared model performance across demographic groups such as women and older patients for each of the 28 datasets to identify when reduced data diversity began to harm predictive quality and when predictive quality was similar for each group. These assessments enabled Alpha to determine the smallest dataset that still met ‘fair’ performance requirements for future AI development (P: *evaluating algorithmic fairness*). Alpha pushed their progressive first-mover approach to open patient data sharing in Germany.

4.1.2. Conditions of use. In Alpha's case, the main setup of the project relied on existing routine data from clinical information systems. Since no patient consent for re–purposing existed, the project operated primarily within the technical dimension of governance by applying strict anonymization procedures. Preventing re–identification became a central concern, particularly when combining large datasets that might inadvertently reveal rare cases or outlier conditions. The initiative therefore implemented iterative privacy evaluations using k-anonymity and other statistical methods to determine acceptable anonymity levels. Common techniques included aggregating age and time indicators, shifting timestamps, and removing quasi-identifiers such as ward names or free-text notes. These measures minimized privacy risks but inevitably led to some loss of granularity in the data. As one medical IT expert summarized, “*When large datasets are combined, the risk to re-identifying outlier cases, for instance, rare diseases, increases*” (Alpha-G).

To balance privacy protection with the analytical needs of AI development, Alpha developed *synthetic data* derived from anonymized patient records (P: *synthesizing data*). The resulting dataset comprised newly generated patient profiles that preserved all features relevant to accurate fall prediction and maintained statistical distributions comparable to those of the original dataset. The team compared different synthetization methods, evaluated metrics for similarity, utility, and privacy, and eventually implemented an autoencoder-based model to generate a synthetic dataset. As an advisory–board member explained,

“ ...we are now constructing relatively realistic models that contain many different components, which come from many different measurements that are put together, ..., so that you practically represent a world in silico like in a flight simulator.” **–** Governance expert (PO-A).

The synthetic dataset became an additional layer of technical governance that enabled data use while preserving patient privacy. By prioritizing data-subject protection, the initiative accepted potential information loss, particularly the underrepresentation of rare cases.

In parallel, Alpha used the *research privilege* provided under Section 9 of the GDPR to justify the secondary use of health data without explicit consent when processing served the public interest (P: *exercising research privilege*). According to the regulation, processing may be permitted without explicit consent under certain conditions, such as for public-health purposes (Recital 54); governance practices need to demonstrate that patient benefits outweigh the remaining risks (PO-A). Following this broad legal interpretation, Alpha gained access to pre-existing datasets. The technical lead explained, “…*To my understanding, we are permitted to use this patient data in accordance with our legal agreements and in line with improving the situation for nursing care*” (Alpha-C). In the absence of explicit consent, data governance structures were reinforced through technological means – synthetic data only and a secure processing environment – to safeguard data subjects' interests.

4.1.3. Data access. A central aspect of Alpha revolved around promoting open access to health data, relying on synthetic data as a key technical safeguard. A challenge emerged in Alpha after consulting anonymization experts: Alpha recognized that although the synthetic data appeared safe today, its governance needed to account for *future attack models* whose methods and technical capabilities were unpredictable. In response, Alpha revised its open data access plan and chose to pursue the plan to place the data on a European research platform, where access would remain subject to a formal governance process. This shift was reinforced by ongoing uncertainty about the legitimacy of openly sharing synthetic data (see Starren et al., 2023). Although stakeholders within the data-providing institution (Clinic 1) broadly agreed that the data were likely safe, none were willing to issue a definitive approval. Alpha engaged ethics committees, data-protection offices, and the open-data institute. In a meeting with the open-data team, one representative stated that the re-identification risk was "extremely low" and that he saw "no issue with publication," yet added, "I also cannot give you a final verdict." Deciding to share patient data, he noted, was being passed along among departments "like a hot potato." (OD-25).

Consequently, the responsibility fell onto the project lead, who had to balance reputational and legal risks against the potential public benefit. To uphold the principle of open knowledge under these constraints, the team decided to initially share their work in alternative ways until a legal opinion was available and further consultations had taken place (P: *opening knowledge*). Instead of publishing the dataset itself, they made their anonymization blueprint and evaluation framework publicly available, together with explanatory Wikipedia articles on the topic. These actions were designed to advance a learning health system. The team held the opinion that even if they could not share the data openly, at least others could learn how they approached it (AD-21). This outcome reflected Alpha's progressive stance: an ambition to maximize transparency and data utility while continuously testing the boundaries of what openness could mean in practice.

4.2. Integrative stance: beta and gamma aligning infrastructures and practices

The data governance approach at *Beta* and *Gamma* was shaped by the goal of *integrating* nursing data into an existing data-sharing infrastructure.

4.2.1. Data scope. Data processors and IT specialists at Beta focused on aligning local data with national interoperability standards. The team decided early to work with a small set of clinically relevant variables, a choice described as allowing "a high degree of interoperability and scalability," since anonymization and mapping efforts increase with each additional field (AD-22). Unlike the data-driven approach of Alpha, Beta consulted nurses to identify which variables were most meaningful in daily practice, and each field was statistically tested for its correlation with fall events. For example, data variables included biomarkers, fall history, and patient compliance. Fluctuations in blood pressure were considered a high-risk factor that had not been prominently discussed in the literature. This pragmatic focus on clinical relevance and standard conformity exemplified Beta's integrative stance, linking frontline knowledge to infrastructural alignment.

At Beta, the team sought to enhance data utility through participatory governance. Workshops and committees involved nurses, data scientists, and IT staff to discuss how the collected data could best support AI development. The focus shifted from "big" to "smart" data collection: instead of maximizing the volume of data, actors limited data collection to features deemed relevant for the specific clinical and research context. Which variables were included was not treated as a purely technical decision, but was negotiated with domain experts. Participants in these formats were given advisory and co-design rights to ensure that decisions about use and access reflected clinical needs.

While Beta initially emphasized minimal data extraction, the team soon revisited this scope as new insights emerged. A nurse explained that "*somehow there is knowledge in institutions about what should be shared*" (Beta-B). Following this intuition, the project added free-text nursing documentation to capture contextual nuances that structured data could not reflect. This process exemplified the project's dynamic re-evaluation of data scope: it minimized redundancy but expanded meaning by integrating user perspectives (P: *data scoping*).

Across Beta and Gamma, such iterative adjustments tied decisions about data inclusion directly to accessibility and governance rights, linking organizational and technical dimensions of integration.

4.2.2. Conditions of use. A challenge emerged at Beta, as the team faced a recurring dilemma about consent procedures. The initial plan involved collecting written consent at the patient's bedside, a narrow legal interpretation of the GDPR. Yet uncertainty grew over whether such consent would remain valid under future regulatory changes. The project manager reflected, "*We came across the fact that different consents are needed to train on an AI model because that is a different purpose than using an AI model*" (Beta-A). Implementing these procedures strained care workflows and risked impeding data availability. As a response, Beta shifted toward a broader interpretation to keep the process manageable, using the research-privilege provision while supplementing it with additional written information for patients and anonymization at the technical level (P: *exercising research privilege*).

Like Beta's initial plan, Gamma adopted a broad consent model that enabled sharing pseudonymized rather than fully anonymized data. However, a participation expert at Gamma made clear that access to pseudonymized data came at the expense of less representative data points, as only a few patients in the hospital were capable and willing to sign the form (Gamma-C). To broaden representativeness, a standardized, multilingual template facilitated this process in the university hospital, even though it demanded additional administrative effort and created more missing data (P: *broadening consent*). Gamma justified its (resource–intensive) stance on data governance with its legal position that "*at the end of the day, consent remains the only legal basis under which we are allowed to share data.*" (PO-E).

Within its federated architecture, Gamma integrated its infrastructure into a clinical repository, storing routine hospital data within a separate area of its hospital information system, and separating the integrated infrastructure from a research repository, where harmonized data could be shared externally through a secure interoperable environment (P: *separating storage*). The team at Gamma adopted a similar integrative stance toward data quality. They favored the state as the coordinator of a federated data infrastructure in which local data hubs retained sovereignty. To ensure data quality and future reusability, nurses and physicians manually reviewed and corrected documentation. As one clinician explained, *"If the algorithm repeatedly receives feedback from the clinicians, 'had no delirium,' this is totally important information to be able to train with it again, because then I know more about whether they were really* *delirious patients"* (Gamma-A). These organizational practices complemented technical measures like centralized error correction and data-hub coordination, balancing the goals of data minimization and usability.

4.2.3. Data access. Data access at Beta was formally defined by its parent organization, which provided standardized templates and procedures. Following the principle that access should vary by storage location and level of anonymity, both Beta and Gamma aimed to make consented patient data available in pseudonymized form to external recipients who completed a formal governance process. Non-consented data, by contrast, had to remain on-site, with only aggregated results allowed to leave the institution. Internal researchers faced a less stringent process (P: *layering access control*). Together, these practices — separating storage, broadening consent, and layering access control — collectively formed a coherent governance stance.

At Gamma, existing local committees were responsible for overseeing requests for data use and access following national standards, while a participatory format reviewed whether relevant stakeholder interests had been integrated. During these discussions, clinicians and nurses concluded that some of the problems the project sought to address were *"primarily interactional, social, or communication problems that could or perhaps even should be tackled without technology"* (Gamma-C). This reflection marked a mature form of integration: recognizing when organizational alignment, rather than technical expansion, achieves the real governance goal.

Together, Beta and Gamma illustrate an *integrative stance toward data governance*: they sought to embed data sharing within existing infrastructures, balance consent with efficiency, and merge technical standardization with participatory reflection. Integration here meant not only connecting systems but aligning expectations, responsibilities, and everyday work practices across institutional boundaries.

4.3. Protective stance: delta safeguarding data sharing

Unlike the other cases, in Delta, the initiative started off with competing industrial stakeholders as potential data providers. Consequently, Delta, relying on private firms, lacked the level of inherent trust (and distrust) associated with public institutions, affecting its stance on governance profoundly.

4.3.1. Data scope. The initial unprocessed data was selected in care homes from ventilators. Technical assessments related to the AI regulation were conducted on separate test data (P: *data scoping*).

"The core idea of the platform is that we don't look at individual data sources in isolation — such as only the data from a medical device or only the data capturing the movement of residents in a building. We want to extract information from a variety of resources." (machine learning expert and project lead – Delta-A).

To maximize the scope of the limited available data, Delta decided to stick with unprocessed data, to 'keep things manageable', and developed their stance on governance around incentivization of data providers and data recipients to join the initiative.

The initiative promised value to voluntary data sharing by providing data quality assessments and statistical analyses (AD-23) on data, for which data providers did not themselves bring up the resources (P: *luring with value*). A potential data provider (Delta-B) outlined the conditions under which they would be willing to share data: "*when you get data from others that is valuable to you, that would be the incentive for us.*"

An illustration of such incentivization is Delta's approach to data quality: Delta's technology–based solution involved automated anomaly detection as one such approach, while, unlike in other cases, the unprocessed data was not further processed. This was only viable since the data was not specifically tied to individuals (but to machines). As one interviewee explained:

"We focus on anomaly detection... We don't change the data, because we don't know what the people who download it want to do with it, but we do want to give an indication" – Project manager (Delta-A).

Even if unprocessed data did not include personal information, it did cover cases with rare conditions, which presented a potential data governance risk. Therefore, Delta's approach aimed at mitigating this risk through a two-fold arrangement linking practices of value luring with the condition that data remain within the provider's own infrastructure. As a data provider from Delta explained:

"[another incentive is] *that you would have the AI model trained locally on your data, for example others would train their AI models on our data via federated learning."* - Medical Technology (Delta-B).

The combination of providing both value and infrastructural assurance functioned as a coordinated approach to govern non-research facilities; furthermore, Delta aimed to maximize data on the provided infrastructure in terms of variable variety, however, minimized the risk of re-identification by limiting its access. In a result report at the end of the project (2025), Delta stated that they had "convinced" a previously unrelated private company and provider of sensor data to integrate their data.

4.3.2. Conditions of use. Although Delta's team chose to provide de-identified data, it adhered to earmarked consent capture, unlike the other initiatives, which either chose broad consent or no consent at all. This decision led to the inclusion of 12 patients in the research project. However, it raised concerns about scaling approaches and the ethical dilemma of obtaining consent from patients connected to ventilators. Considering a narrow legal application, the conflict with limited consent capture due to patients' physical conditions has remained (yet) unresolved, particularly when considering scalability. If scalability is compromised, the project may need to limit data collection to selected, competent patients.

Delta's stance executed a narrow interpretation of legal conditions at the expense of scalability, which was also demonstrated in their data pipeline. The initiative ensured GDPR compliance and privacy considerations of data subjects with media disruption, using USB sticks to transfer the data once a week safely, yet manually, to the secure data processing environment. With this architectural choice, scalability was again compromised in favor of mitigating risks to data subjects' interests. As normative governance conflicts, i.

e., how to ensure safe data transfers in compliance with GDPR, are unfit to be solved with the technological dimension of governance, i. e., a secure processing environment, an organizational solution, namely human resources, serves as a substitute.

4.3.3. Data access. Technically, the Delta team adopted a fully centralized approach. Discussions have included the consideration of a data safe haven model, which brings in a trusted third party in key governance issues. Regarding data sharing and access, these initiatives employ integrated technical, organizational, and normative governance constructs to manage legal uncertainties. One example is the close collaboration and initial clarification of public-private partnership models at Delta (DSE-A). Delta always aimed to find a private platform owner (Delta-A; Delta-E). The absence of a public ethics committee they could directly turn to and the lack of dedicated organizational structures exacerbated this challenge. Our research identified the conceptualization of complementary patient boards alongside scientific ones, granting affected group representatives veto rights or advisory roles; secondly, adherence to legal requirements focusing on scientific principles and safeguarding public interests. External representatives of these boards confirmed these principles. In contrast to these conceptualizations, Delta's team did not make the data accessible to others outside the consortium in the project duration. Yet, it mentioned it would publish results in the form of both articles and as code, which, however, was not realized in the project duration (P. *opening knowledge*). Their approach, in principle, aimed to establish a community model (e. g., a "rather open" approach to data, Technical Lead, Delta).

In Delta's design for data sharing and access, their main practice – to lure with value – became present again: Operators were incentivized through a feedback system on the results of data analysis, which was expected to provide additional value.

> "It would be exciting if I could also use the platform when the AI developer has developed his solution with it, that I can perhaps make my AI model available again as a starting point for further development by another team perhaps I have the opportunity to offer a license on how to continue using my model."
>
> Delta-E

Table 5

Stances toward data governance in data sharing for AI development.

	Progressive	Integrative	Protective
Areas of Tension			
Data Scope (Min - Max)	Large volumes of unprocessed data, then anonymized, later little synthetic data is created	Mid-range variety and sample size, depending on anonymized and pseudonymized data set	Large volumes of high frequency data from a few samples
Conditions of Use (Narrow -Broad)	Broadest legal application with research privilege, no consent	Broad future use only when consent obtained	Narrow legal application, specific consent
Data Access (Closed -Open)	Open to wider audiences	Controlled – for community	Restricted to network partners
Complementary Data Governance Practices			
To Address Data Scoping	DS: Adaptive creation of specific, tailored data set versions for modeling and AI useǂ LV: Quick insights and data quality improvements for providersռ EF: Providing verifiable reports on fairness as legal accountabilityǂ	DS: Participatory data scoping with domain participantsǂ LV: Participatory workshops explaining the value of harmonizing with parentռ CQ: Consider data provenance; cleanse unstructured dataǂ EF: If data is pseudonymized, only model sharing, not data sharing. Evaluation of models is necessaryǂ	LV: Novel explanations how data interlinkages will increase data provider valueռ CQ: Consider metadata such as user data from data providers; mark data quality as commentary on unprocessed dataǂ EF: If data is unprocessed, evaluating models gains relevance; integrate model outcomes as reinforced learning in dataǂ
To Address Conditions of Use	SD: Use of emerging, privacy enhancing mechanisms (synthetic data)ǂ RP: Evaluation of common good criteria for data usage; Prohibition of re-identification as contractual conditionռ	LP: Adaptation of governance standards from parent organizationԾ BC: If data is shared, consent must be obtained. Broadening consent allows for variety in use casesռ SS: physical separation between identifiable and anonymized datasetsǂ	BC: Working only with earmarked consent, if data is shared; no consent if data remains on premise and model is integrated ռ
To Address Data Access	OK: Opening access to aggregated data results (e.g., blog posts, blueprints, code)ռ LC: Pseudonymization by a trusted third party; identification of data usersԾ	EC: Use of organizational committees and communities to decide about data accessԾ LC: Provision of data per request on parent platformԾ	OK: Sharing code and other results as compensation to use data ռ LC: Provide access to the platform for affiliated data users onlyԾ
Mode of Integration to Infrastructure	Build own decentralized open governance network and plugin into open science infrastructures	Integration of decentra-lized providers to central parent organization	Centralized closed governance network
Case example	Followed by Alpha	Followed by Beta and Gamma	Followed by Delta

Legend of practices: Data Scoping (DS), Luring with Value (LV), Broadening Consent (BC), Exercising Research Privilege (RP), Synthesizing Data (SD), Contextualizing Data Quality (CQ), Separating Storage (SS), Leveraging Parents (LP), Opening Knowledge (OK), Evaluating Algorithmic Fairness (EF), Engaging Communities (EC), Layering Access Control (LC).

Symbol coding of primary dimensionality: ǂ technical, Ծ organizational, ռ normative.

Delta exemplified a *protective stance* toward data governance by emphasizing security and control over openness. Its centralized infrastructure and manual data transfers minimized data access and maximized heterogeneity of the data in place; however, scalability of the model was limited.

5. Discussion

5.1. Main findings

This paper examined how data governance is enacted as practices to reconcile unique tensions in data sharing for AI development. Through a *multiple case study* on data-sharing initiatives in the context of a public funding effort in AI for nursing in Germany, we have derived three stances with twelve data governance practices that aim to balance out stakeholder interests across three core areas of tension.

A central insight from our study is that certain requirements when sharing data for AI development are inherently in tension and cannot be fully met simultaneously. Data governance in multi-stakeholder AI projects is essentially a process of reconciling these incompatible interests: (i) *Data Scope* considers the scope and integration of data sets (data minimization versus representative, large data sets) for reasons of privacy and potentially greater analytical value; (ii) *Conditions of Use* considers the application of regulatory requirements (narrow vs. broad) for legal assurance on repurposing, consent, data storage, and broad usage, and (iii) *Data Access* considers accessibility rights as a tension between open data for AI development versus control and autonomy. We found that these tensions were addressed by each of the four projects, although each with a different emphasis.

5.2. Stances toward data governance in data-sharing initiatives

We summarize our findings in three distinct stances, which we have termed *Progressive*, *Integrative*, and *Protective.* The Progressive stance is characterized by a large, anonymized data set used within the initiative, of which only a portion was aimed for sharing after synthetization; it is also characterized by wide use opportunities due to reliance on the research privilege, and finally, more *open* data access. The Protective stance, in contrast, entails closed or highly gated access, more expansive data scope but on a restricted sample, and broad reuse of data but only internally or for approved actors. The Integrative stance falls in between on each dimension – it features selectively controlled access, a moderate level of data integration, and conditional reuse options that are negotiated among participants. We characterize the practices of a stance as those that directly address its primary favored side of the conflict and consider them complementary when they mitigate the resulting trade-offs (see also Appendix 6). Table 5 provides a comprehensive overview of the main characteristics with respect to areas of tension and complementary data governance practices. In the following, we discuss each stance in turn, clarifying which tensions it prioritizes and how its practices cope with those tensions.

The *Progressive* data governance stance copes with tensions by prioritizing the technological dimension and by leveraging open standards (e.g., decentralized storage in secure systems or the usage of privacy-enhancing technologies). Starting from an ambitious but ambiguous vision to create a decentralized, community-driven governance network, the stance stabilized by emphasizing open sharing via providing representative synthetic data. As a result, the stance involved complementary data governance practices, such as the use of data without explicit consent in accordance with Section 9 of the GDPR for research purposes. It also requires outlined accountability for the use and access committees, and iterative, multi-purpose data analytics to incentivize data sharing by data providers (e.g., reinforcement of organizational innovation). Alpha's stance followed this approach.

From a theoretical standpoint, the ***Progressive*** stance demonstrates how technical innovation is entwined with normative legitimacy in practice. In prior work, open data initiatives often emphasize policy mandates (Brous et al., 2016; Cao & Iansiti, 2021) or the availability of platforms. The results of this study are in line with research advocacy at the intersection between open data and fairness (Seastedt et al., 2022) and show that opening large sets of sensitive data for AI requires a finely tuned, orchestrated set of (privacy-preserving) practices. The Progressive stance thus also expands on the technical mode of data governance identified by Winter and Davidson (2019), who concentrate on ensuring contextual integrity of data illustrating that a technology-driven approach in data governance is not purely technical, but relies on a network of complementary practices. Notably, Paparova et al. (2023) similarly found that when multiple actors enact data governance, emergent dynamics require adjusting practices on the fly.

The ***Integrative*** data governance stance resolves tensions by prioritizing the organizational dimension, adapting governance standards from parent organizations (e.g., consortia guidelines, consent templates). Its *hierarchical* vision is a centralized governance network with decentralized accountability, making data access restrictive to community members and its use identifiable. As a result, this stance involves complementary practices, such as separating data storage, which goes hand in hand with a consent framework involving organizational and technical steps, and transferring the costs of formal legal vetting to the parent organization to incentivize data providers. With variations in the employed technological architecture (centralized versus decentralized), Beta and Gamma followed this model, resonating with (Susha et al., 2023, p. 1) finding that to an extent "the partnership model determines the types of drivers and challenges to sharing data in a partnership."

By showing that bridging organizations in data sharing is a possibility, the stance expands on the idea of data collaboratives in the literature (Susha et al., 2023), but it also cautions that the practice of governance in such collaboratives requires ongoing effort to balance local and central interests. By bringing diverse stakeholders together, this stance builds what has been termed in a white paper by the European Commission an "ecosystem of trust" (European Commission, 2021, p. 3) for data sharing – a socio-technical environment in which participants feel confident that data will be used for agreed-upon purposes and that individual rights are safeguarded even as collective value is generated. The Integrative stance strikes a balance between progressive and protective approaches, aiming to get the benefits of data synergy while ensuring the safeguard of collective governance. This approach aligns with the notion of "data space" governance (e.g., demonstrated by Costabile & Øvrelid, 2023). As Kilgus et al. (2024, p. 511) observe, data spaces, such as the European Health Data Space, "build on participatory governance mechanisms and are based on federated infrastructures," enabling actors to exchange and reuse data under "jointly established values, goals, and standards." The stance showcases the value of an orchestrating entity or steering committee that can balance interests – similar to the notion of a "legitimized orchestrator" in data spaces.

The ***Protective*** data governance stance is an industry-driven, standardized approach to building a data-sharing infrastructure. This normative-driven approach relies on multiple data sources as input, addressing conflicting interests with trust-building structures and incentives, such as shared data outputs (e.g., sharing machine learning models), while governance is continuously negotiated among stakeholders. Its *exclusive* vision is a centralized, closed governance network that allows data access according to affiliations. As a result, this approach involves aligning governance practices, such as participative integration of affected stakeholders (e.g., through earmarked consent or multi-stakeholder use and access committees) and the provision of unprocessed data to minimize effort and incentivize data holders. Gamma followed this stance.

The implications of this stance can, on one hand, lead to quick innovation because the richness of data and freedom of internal reuse give AI developers a strong advantage. On the other hand, it may raise concerns about equity and transparency. When valuable data is siloed, only a narrow set of actors benefit, and broader progress may be stifled. Winter and Davidson (2019) note that governance choices often result in outcomes where either a specific agent captures most of the value or value is broadly shared – with the protective stance skewing toward the former. A prominent example is the collaboration between DeepMind and a UK hospital, which provided millions of patient records to a private company resulting in backlash over privacy violations (Winter & Davidson, 2019) and one-sided value capture (Powles & Hodson, 2017). The Protective stance represents an "innovation behind closed doors" approach that may be effective in the short term, but potentially unsustainable if public and regulatory pressures demand greater openness.

In essence, the three stances differ mainly in how they balance openness, coordination, and control: Progressive governance emphasizes open access, synthetic or anonymized data, and broad reuse; Integrative governance relies on shared organizational rules and selective access to balance competing interests; and Protective governance limits access and reuse to trusted actors in order to maximize oversight and minimize risk.

There are notable commonalities in stances toward data governance. The most common issue in data scoping is navigating the conundrum between collecting large sets of heterogeneous data for broad use, favored by data engineers and ML scientists, and the demand for specific, hypothesis-driven use by data protection personnel, clinical trial offices, biostatisticians, and sometimes clinical and nursing scientists. This challenge mirrors the conflict between data-driven and theory-driven research identified by Maass et al. (2018) and is often not resolved synergistically, as proposed by the authors, but instead leads to unsustainable solutions favoring one over the other. In doing so, this work empirically substantiates the often-cited tension between enabling open data innovation and ensuring privacy. For example, Winter and Davidson (2022) describe how disparate policies and goals for data protection versus data innovation may hinder socially desirable innovation if not harmonized, for example, when only focusing on privacy-enhancing measures. Our findings resonate with the idea that governance choices ultimately determine the allocation of *value and risk*: A *protective stance* concentrates both value and control within a few trusted entities, minimizing exposure but limiting shared benefits. A *progressive stance* expands access and potential value creation but increases the need for safeguards against misuse. An *integrative stance* seeks balance, coordinating rules and responsibilities to distribute value and risk more equitably among participants.

A commonality in data usage conditions is finding an appropriate level of anonymization, balancing data privacy concerns with information utility. This work empirically substantiates the often-cited tension between enabling open data innovation and ensuring privacy. In this area of data governance, technological dimensions of governance gain relevance. While technological solutions (Kearns & Roth, 2019) can help, complementary organizational and normative practices are necessary before any data flows. As Winter and Davidson (2019) note, and as Khurshid et al. (2020, p. 63) succinctly state, "health data moves at the speed of trust." Therefore, trust-enhancing measures are emphasized in all data governance stances.

A final commonality is the conflict between providing large datasets for open communities and maintaining controlled, secure access. In this conflict, organizational governance through committees, review boards and other structural arrangements can become central. This is akin to the conundrum discussed by Davidson et al. (2023) regarding the exposure of open data to digital innovation.

5.3. Implications

Much of the data governance literature conceptualizes governance through formal decision rights, policies, principles, and accountability structures (e.g., Abraham et al., 2019; Brous et al., 2016; Khatri & Brown, 2010). Our findings do not contradict this work but show that it is incomplete for AI-oriented data sharing. In our cases, governance became actionable only when such formal arrangements were enacted through sociotechnical practices that jointly reconciled tensions around data scope, conditions of use, and access. Theoretically, this shifts attention from governance as a designed structure to governance as a situated accomplishment.

Our findings demonstrate that contemporary stances toward data governance are formed by patterned bundles of sociotechnical practices rather than singular policies that attempt to navigate these tensions. We conceptualize each stance as a distinct convergence of practices that jointly reconcile conflicting interests around data sharing. These practices are not isolated; they work together. Each stance combines practices with complementary practices that compensate for their trade-offs. This means that one type of stakeholder or expertise often takes leadership in implementing a governance stance, while others provide critical support. This situated perspective allows for seeing data governance not just as formal rules or structures (Brous et al., 2016), but as sets of practices that collectively work to reinforce or alleviate the central tensions.

Data Governance Practices for Data Sharing Reconcile Conflicting Interests.

The first implication of our study is to conceive data governance as an emergent endeavor of reconciling conflicting interests in data; departing from a conventional view, in which data governance has been seen as a means for ensuring data protection (e.g., Khatri & Brown, 2010).

Recent contributions have shown that data governance requires bringing together and reconciling the requirements of multiple stakeholders (e.g., Burmann et al., 2023; Paparova et al., 2023). In line with such work, our study presents concrete practices that data-sharing initiatives in nursing care use to perform reconciliation work. As we argue, this conceptual synthesis extends existing contributions on governance practices, such as patterns of practices by Mikalsen and Monteiro (2021). Practice-based studies such as Parmiggiani et al. (2022) show how data are made analytically workable in the often-disguised preparation work of data science. Our concern is different: we explain how data sharing becomes governable across organizational boundaries when actors must justify trade-offs among privacy, control, and future reuse. This shifts the focus of the analysis from data work to governance work, emphasizing the growing need to extend beyond organizational boundaries and adopt a participatory approach (e.g., Paparova et al., 2023; Safadi et al., 2024). Our contribution further adds and refines empirically identified practices from prior IS or management research such as ways to *synthesize data* (Arora et al., 2023), which was specifically used in one of our cases to make data shareable; *contextualize data quality*, an epistemic practice to increase the value of data by providing valuable contextual information (see Günther et al., 2017; Patecka et al., 2025); and *leveraging parents*, an integrative practice highlighting the emerging importance of parent organizations such as data sharing consortia (see Gray et al., 2022).

While we acknowledge that our data collection was limited time span, it is similar to Barley's (1986) study in the sense that it sees structuring actions emerging sequentially in response to specific challenges over time. For instance, while issues of data scoping and usage conditions had been present from the start, issues of sharing and access emerged later when the first pilot use cases or access requests arrived, despite having been long anticipated.

5.3.1. Data governance practices require orientation

The second implication of our study is that distinct data governance stances emerge as “option repertoires” (Noesgaard et al., 2023) from the way normative, organizational, and technical practices are combined. We find that these practices work together rather than independently, forming an integrative logic that links different governance dimensions into coherent orientations. Whereas prior studies have often examined the legal (Shabani, 2022), organizational (Abraham et al., 2019), and technical (Otto et al., 2022) dimensions of data governance separately – or have treated them as analytically distinct yet interlinked *forms* of data governance (Winter & Davidson, 2019) – our analysis offers a different perspective by showing how multiple, interacting practices become informed by broader orientations, which we identify as progressive, integrative, and protective stances toward data governance.

This implication follows from the situated view, as we illustrate the utility of studying data governance in situ. Our finding is that data governance is malleable. Based on initial directions and imprints, it can take different directions, even if the project is funded in the same funding line and brings together similar institutions such as clinics, AI researchers, and startups. Malleability, in our account, therefore refers to bounded but consequential variation in how governance is enacted under similar conditions.

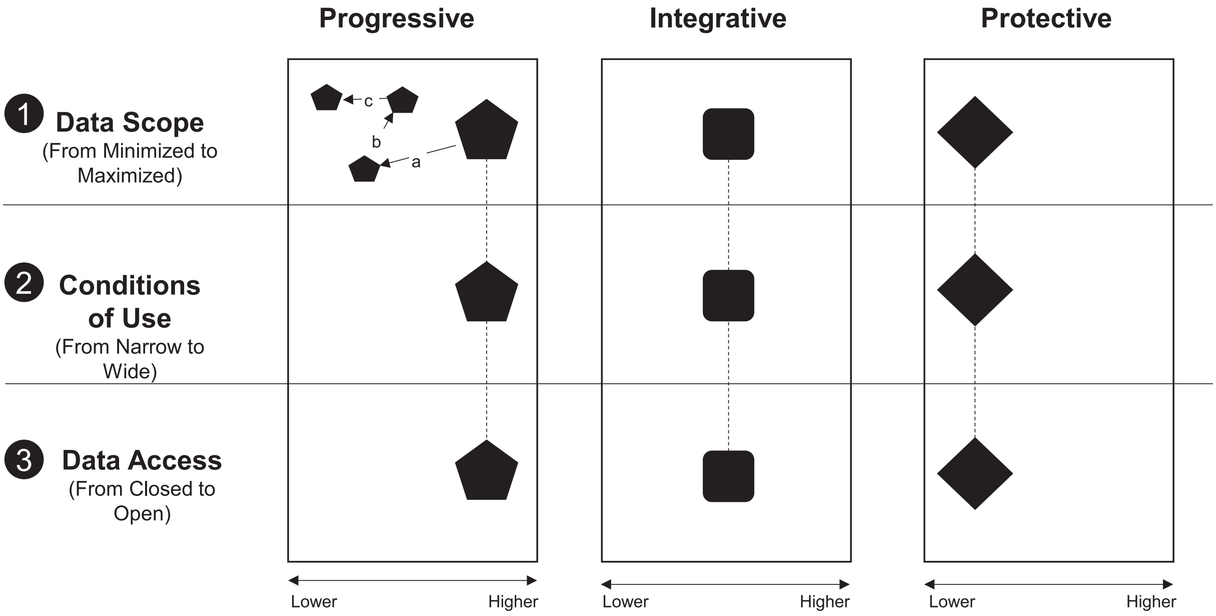
Similarly to our study, Paparova et al. (2023) considered data governance practices as emergent and showed that complex dynamics unfold when multiple actors are enacting data governance practices. Extending this work, we introduce the idea of data governance practices as *guided*: technical and organizational data governance practices require orientation in order to work together with normative practices in ways that depend on the contextual conditions and unfolding of the dynamics of a specific setting. This was exemplified in the case of the *Progressive* stance, where data is shared with no consent, considering additional practices such as verifiable and well-documented anonymity assessments. Gray et al. (2022) also mention the need for technological alignment with the “Right to Science”/research privilege framing; our study illustrates in detail how inherent linkages between practices are informed by stances.

5.3.2. Data governance for future AI development needs to consider data-related risks and benefits prior to the actual AI development

The last implication of our study is that data governance for AI development needs to consider data-related risks *and* benefits (e.g., minimal vs. representative data scope, narrow vs. broad legal application, open vs. controlled access). In this study, we view data governance for AI development as a reconciliation process that expands the current scope of governance in response to the requirements of the GDPR as well as emerging legal regulations such as the EU AI Act – and is thus a response to the work by Davidson et al. (2023) calling for work on new forms of data governance. Here, the new form integrates AI-related risk and benefit management that has been underdeveloped in traditional governance models. Examples of this expansion include transparency obligations related to downstream data outcomes concerning bias and fairness metrics, as well as considerations of data utility such as the volume of data and the interlinkages of heterogeneous datasets (Hillebrand et al., 2023). The focus on data governance for general reuse (e.g., for research innovation) gives way to the specification of a governance term that clarifies reuse for AI development, intensifying current positions in IS literature on the tension of data governance for digital innovation (Vial, 2023). While Frühwirth et al. (2024) emphasize risks in the context of AI development,[3] our perspective includes both data-related risks and benefits as means for governance.

---

[3] This issue of mandating transparency for algorithms is increasingly prominent in data governance debates. For example, Mühlhoff and Ruschemeier (2024) warn that one of the greatest data protection risk in the context of AI is the *abusive secondary use of trained models*, which is a risk largely unchecked by current GDPR or AI Act rules. Our case illustrates early attempts to confront this risk via governance: by treating trained models and derived insights as in-scope for oversight and not just the unprocessed data.



a – Data scope is minimized via deletion and selection, b – data scope is increased by adding new use cases/data
c – Data is branched and data scope is further minimized through privacy-aware synthetization

Fig. 1. Data governance stances in areas of tensions.

This study highlights additional contextual data governance practices for AI development, particularly addressing the fundamental tensions between data utility for machine learning and privacy/regulatory compliance. One example is *evaluating algorithmic fairness,* which illustrates the negotiation and review process to transparently clarify divergent perspectives on training data and model outcomes for potential biases, demographic representation, and fairness metrics.[4] This practice extends beyond traditional data quality measures to address AI-specific concerns such as dataset shift, label quality, and class imbalances that can lead to discriminatory model behavior (e.g., Bauer et al., 2024). For further elaboration, we consider the convergence of different practices and their implications for AI development, particularly the need for high-quality training data at scale, shown in Fig. 1.

Different governance stances prioritize different characteristics that are critical to AI development, such as data scope. Each approach is architecturally designed to balance these characteristics, resulting in trade-offs in the AI development pipeline.

The *Progressive* stance facilitates the sharing of large datasets through privacy-preserving techniques and the use of synthetic data generation. While this approach supports the data volume requirements of deep learning models, it can introduce data robustness issues that impact model performance (see Hernández et al., 2022), particularly for rare edge cases or complex feature interactions. In Fig. 1, we see how the scope of data may shift over time via different transformations as a result of ongoing negotiations to mitigate tensions.

The *Protective* stance, in contrast, preserves data integrity by allowing access to unprocessed data, which is particularly valuable for developing precise medical AI applications (see Starren et al., 2023). However, the strict consent requirements significantly constrain access to data, potentially limiting the model's utility for AI research.

The *Integrative* stance attempts to balance these competing demands by distributing accountability and availability of shareable data along diversified consent measures. While this enables broader data access while preserving privacy, the organizational overhead and technical complexity of distributed data storage may impede rapid experimentation and iteration cycles typical in AI development. It may call for “methods and semi-automated tools to facilitate data work across organizational entities and distributed datasets” (Jakubik et al., 2024, p. 508). This stance also faces challenges in maintaining consistent data quality and standardization across different organizational contexts. Taken together, a picture emerges that reinforces that a data governance stance for AI use is not a monolithic concept, but rather a collection of coordinated practices aimed at reconciling particular tensions in data use.

For practitioners, the three stances should not be read as fixed templates, but as repertoires for navigating recurrent tensions in healthcare data sharing for AI development. Because these tensions are interdependent and ongoing rather than problems that can be solved once and for all, governance requires selective prioritization combined with complementary mitigation. When openness is prioritized, as in the Progressive stance, additional safeguards become critical, including privacy-preserving transformations, documented fairness checks, and formal review procedures for access. When coordination with existing infrastructures is prioritized, as in the Integrative stance, managers should invest in consent design, separation of storage, layered access rules, and participatory

---

[4] Fairness metrics consider performance outcomes such as precision, recall, accuracy for different subgroups (in relation to age, gender, health status etc.)

alignment with clinical users and parent organizations. When control is prioritized, as in the Protective stance, managers need to recognize the resulting limits in scale, external legitimacy, and collective learning, and compensate through clear value propositions and tightly defined access arrangements.

Our findings also inform choices among legal and organizational arrangements for reuse. Earmarked consent can increase perceived legitimacy and provider control, but it may sharply constrain scale and representativeness. Broad consent can support wider reuse, but it requires organizational investment in communication, storage design, and access governance. Reliance on research privilege may enable secondary use without repeated consent, but only when accompanied by strong justificatory and technical safeguards.

6. Conclusion

This study has considered the intricate complexities of data governance of data-sharing initiatives for future AI development. Our examination of four cases indicates that, although risk-averse sharing cultures in healthcare and strict data protection laws present challenges across cases, different stances toward data governance for sharing data emerge – Progressive, Integrative, and Protective. These represent alternative overall orientations to navigate these intricacies. These stances explain how different practices across normative, organizational, and technical dimensions become aligned to tackle the major issues related to “making data available” for future AI development in healthcare. Yet, while these stances have the potential to enhance collaboration across stakeholders, their effectiveness is not without limitations. The study points out the hurdles in implementing data governance effectively and suggests that the real world makes data governance hard work. By discussing stances toward data governance in the German context that aim to mitigate issues in data sharing, this research contributes insights into the ongoing effort to cultivate a culture of innovation within data-driven healthcare environments, adding to insights from other countries (e.g., Benfeldt & Persson, 2025; Paparova et al., 2023; Winter & Davidson, 2019). However, it also highlights the need for more situated data governance frameworks that can better reflect the dynamic needs of data-sharing initiatives for AI development, emphasizing that the path to enhancing healthcare outcomes and fostering collaborative innovation in nursing care is fraught with complexity and ongoing challenges.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used ChatGPT, Hemmingway Editor, Claude, and Copilot in order to check for spelling and grammar mistakes, to improve the readability of the text, and for enhancement of understandability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

CRediT authorship contribution statement

Rahel Gubser: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Matthias Schulte-Althoff:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jörg Pohle:** Writing – review & editing, Validation, Resources, Project administration, Methodology, Investigation, Conceptualization. **Nils Heinemann:** Writing – review & editing, Project administration, Investigation, Data curation. **Felix Balzer:** Writing – review & editing, Supervision, Project administration, Methodology. **Daniel Fürstenau:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Author statement

Rahel Gubser: Conceptualization, Data collection, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing - review & editing.

Matthias Schulte-Althoff: Data collection, Data curation, Investigation, Methodology, Project administration, Writing - review & editing.

Jörg Pohle: Investigation, Visualization, Methodology, Writing - review & editing.

Nils Heinemann: Data collection, Visualization, Writing - review & editing.

Felix Balzer: Funding Acquisition, Supervision, Resources, Writing - review & editing.

Daniel Fürstenau: Conceptualization, Data collection, Data curation, Formal analysis, Funding Acquisition, Investigation, Methodology, Supervision, Resources, Writing – original draft, Writing - review & editing.

Acknowledgements

We thank the members of the Freie Universität Berlin, IT University in Copenhagen, and the Institute of Medical Informatics at Charité - Universitätsmedizin Berlin for their valuable insights on earlier drafts of the paper. We are also grateful to the project teams for granting us the opportunity to conduct the study. Further thanks go to Elizabeth J. Davidson, Aleski Aaltonen, Petros Chamakiotis, and Jonathan Wareham for commenting on earlier versions of the paper. This study received funding from the German Federal Ministry of Research, Technology and Space under the ProKIP project (BMFTR grant number 16SV8836) and the KIP-SDM project

(BMFTR grant number 16SV8854). The first author received a scholarship from the Ernst Ludwig Ehrlich Studienwerk, which supported her independent research position while enabling first-hand access to data and insights.

Appendix 1: Overview of data governance practices in IS literature

Area of Tension	Practice Domain	Practice	Literature
Data Scope	Data Selection	Setting up appropriate contracts[D]; data sharing agreements[D] / Service Level Agreements; Data Processing Agreement[D]	Frühwirth et al. (2024) Abraham et al. (2019) Paparova et al. (2023) Winter and Davidson (2022)
		Segregating data for different purposes[R]	Paparova et al. (2023)
		Compensate for data sharing[R]	Gray et al. (2022)
	Collaborative Value Definition	Patient engagement workshops	Winter and Davidson (2022)
		Articulate value proposition[D]	Davidson et al. (2023)
		Synchronize expectations[D]	von Grafenstein (2022)
		Articulate implicit norms[D]	Hofmann et al. (2017)
		Aligning the technological policies with the Right to-Science framing[R]	Gray et al. (2022) von Grafenstein (2022)
		Coupling data with real world phenomena	Benfeldt and Persson (2025)
	Data Quality	Managing data as an enterprise asset	Benfeldt Nielsen (2017)
		Curating data-based impressions (clicking pages in the learning to appear busy)	Aaltonen and Stelmaszak (2023)
		Define metrics for data quality[R]	Paparova et al. (2023)
		Audit algorithms[R]	Celiktutan et al. (2024)
		Classify data sources[R]	Frühwirth et al. (2024)
Conditions of Use	Consent	Establish opt-out of data reuse	Winter and Davidson (2022)
		Introduce broad consent[R]	Gray et al. (2022)
		Obtain consent for remote monitoring[R]	Paparova et al. (2023)
	Data Security, Privacy & Control	Build trusted relationships, integrating trustful third parties[R]	Frühwirth et al. (2024) Gray et al. (2022)
		Define business owners of data assets clearly[R];	Khatri and Brown (2010)
		(De-)centralizing governance and accountability[R]	Aaltonen et al. (2023)
		Create compartmentalized "Unitary Spaces" for individual negotiation[D]	Cepa and Schildt (2023)
		Integrate bilateral review system	Chen et al. (2024)
		Share models instead of data[R]	Frühwirth et al. (2024)
		Automated tracking and performative reactions as consequences	Frühwirth et al. (2024) Aaltonen and Stelmaszak (2023)
		Use privacy-preserving, secure devices for storage[R]	Parmiggiani and Grisot (2020)
		Apply federated learning and other privacy preserving data handling[R]	Frühwirth et al. (2024) Parmiggiani and Grisot (2020)
		Share only synthetic data; use anonymized data[R]	Frühwirth et al. (2024)
		Find a service on-premise solution[R]	Frühwirth et al. (2024)
Data Use & Access	Access	Give role-based access[R]	Dyke et al. (2018)
		Deny access for specific organizations (e.g., Google)	Winter and Davidson (2022)
		Validate with Triple "A" Registration process (Authentication, Attestation, and Authorization)	Dyke et al. (2018)
		Establish governmental review board to provide ethical oversight and grant access	Winter and Davidson (2022)
		Establish an Independent Review Panel	Winter and Davidson (2022)
		Build communication practices (training)[D]	Abraham et al. (2019); Parmiggiani and Grisot (2020)
		Collaborate with community (e.g., through moderators) in participatory way to gain needs, standardize, and educate users	Safadi et al. (2024)
		Facilitate multi-stakeholder dialogue forum[D]	Hofmann et al. (2017)
	Use	Create unalterable records of data access and use[D]	Parmiggiani and Grisot (2020) Gray et al. (2022)
		Design data governance serviced-based (i.e. designing APIs with embedded rules and usage quotas)[R]	Vial (2023)
		Create unalterable records of data access / use (immutable audit trails)	Gray et al. (2022)
		Document the past (data provenance)	Jarvenpaa and Essén (2023)
		Curate data zones for audited, production use[R]	Vial (2023)
		Include multi-stakeholder governance board[R]	Paparova et al. (2023)

[R] Practice is mentioned / discussed in Results.
[D] Practice is mentioned in Data, yet not further discussed in Results.

Appendix 2: Summary of data sources

Source	Type of Data	Use of Data in Analysis (Phase 1–5)
Semi-Structured Interviews	- **44** interviews conducted between October 2022 and April 2025 - **1743** min of audio in total - **25** distinct representatives of #Cases: project managers of developing health data-sharing initiatives; **6** representatives of #ParentOrganizations: data sharing infrastructure builders and representatives from public initiatives, with expertise in science coordination, interoperability, governance and anonymity; **6** representatives of #DataSharingEnablers with expertise in economics, infrastructure building in public and private areas	Phase 1 – Case exploration - Case Descriptions (Table 1) - Open Codes on vision, functionality, economic model, stakeholder responsibility Phase 2 – Expanded data collection and practice coding: - Open coding following Barley (1986) - Phase 4 – Linking practices to approaches - Axial Codes on dimensions within practices - Axial Codes on consequential interrelationships of different practices - Selective Codes on 3 main data governance approaches; (Table 3)
Archival Data	- 202 pages of documents of anchor cases consisting of 4 project plans, anticipated architectural designs, information on roles and stakeholders (*#Cases*)	Phase 1 – Case exploration - Case Descriptions (Table 1) Phase 2 – Expanded data collection and practice coding: - Open coding following Barley (1986)
	- 134 pages of technological guidelines and templates consisting of data privacy guidelines, anonymity guidelines, consent templates (*#ParentOrganization*)	Phase 5 – Reconceptualizing core tensions: - Challenge codes, compare memos for analytical rigor and shared interpretation
Field Observations	- 4 participation workshops with users for data-driven nursing use cases - 4 workshops with *#Cases* participants - 2 days of nursing data convention (*#Cases*)	Phase 4 – Linking practices to approaches - Axial Codes on governance dimensions within practices - Axial Codes on interrelationship between different practices - Selective Codes on 3 main data governance approaches; (Table 3)
	- 1 conference on secondary data use with industry and academia (*#DataSharingEnablers*) - 2 conferences on medical data sharing (*#ParentOrganization*)	Phase 1 – Case exploration - Particularly on existing parent organizations, relevant stakeholders and their contribution to data sharing governance
	- 1 public hearing to legal referendum on German health data usability act	Phase 1 – Case exploration - Current regulatory environment and getting familiar with key stakeholders in the German data sharing ecosystem
	- 300 MB of data (weekly jour fixes, workshops, technical discussions, etc.), notes, presentations, and further documents (e.g., study plans, work package progress reports, weekly meeting notes, data protection concepts)) from the Alpha #Case	Phase 2 – Expanded data collection and practice coding: - Open coding following Barley (1986)

Appendix 3: Interviewee list

			Interview Date			
Participant	Gender	Field of Expertise	Oct—Nov 22	Feb—Mar 23	Nov—Dec 23	Jan 24—Apr 25
Cases of Data Sharing for Nursing AI						
Alpha-A	M	Medical Technology, CTO	X			
Alpha-B	M	Nursing Science	X			
Alpha-C	M	Machine Learning		X	X	
Alpha-D	F	Social Economy		X		
Alpha-E	M	Clinic Management, CEO		X		
Alpha-F	F	Nursing Science		X		
Alpha-G	M	Medical IT Machine Learning				X
Beta-A	F	Medical IT		X	X	
Beta-B	F	Nursing		X	X	
Beta-C	F	Medical IT				X
Gamma-A	M	Clinical Medicine, Medical IT	X		X	
Gamma-B	M	Nursing Science	X			
Gamma-C	M	Participation			X	

(*continued on next page*)

(*continued*)

			Interview Date			
Participant	Gender	Field of Expertise	Oct—Nov 22	Feb—Mar 23	Nov—Dec 23	Jan 24—Apr 25
Gamma-D	M	Responsible AI			X	
Gamma-E	F	Medical IT		X		
Gamma-F	M	Medical IT		X		
Gamma-G	F	Medical IT		X		
Gamma-H	M	Medical IT		X		
Gamma-I	M	Medical IT		X		
Gamma-J	F	Medical IT		X		
Delta-A	M	Medical IT	X			
Delta-B	M	Medical Technology		X		
Delta-C	F	Nursing Science		X		
Delta-D	M	Nursing Science		X		
Delta-E	M	Medical IT	X		X	
Parent Organization						
PO-A	F	Clinical Medicine, Medical IT, Governance		X	X	
PO-B	M	Medical IT		X		
PO-C	M	Medical Technology		X		
PO-D	F	Medical IT, Governance		X		
PO-E	M	Medical IT, Governance		X	X	
PO-F	M	Medical IT, Interoperability		X		
Data Sharing Enabler						
DSE-A	M	Data Sharing Expert, CEO			X	
DSE-B	M	IT			X	
DSE-C	F	Economics		X		
DSE-D	M	Medical IT		X		
DSE-E	M	Clinical Medicine, CEO			X	
Dropout	M	Clinical Medicine		X		

Appendix 4: Interview guides

#1 Interview guide ($n = 6$)

The first interview guide was used for interviewing participants from the four anchor cases. It helped us understand anticipated architectural choices regarding data infrastructure and thus develop more detailed interview guides for the following interviews.

General Background

- What is your educational and professional background?
- What is your role and responsibilities within the organization?
- How does your daily routine look like?

Anticipated Technical Solutions

- What is the goal and anticipated outcome of your project?
- What type of data do you process? What are the data sources in your infrastructure?
- What is your choice of architecture for the infrastructure? (storage, analysis)
- How are decision making processes structured?
- Who are the infrastructure's stakeholders?
- What is the stakeholder engagement regarding privacy and patient consent?

Economic Sustainability

- Whom are you planning to appoint as the operator of the data infrastructure?
- What are your intentions for the post-project period?

#2 Interview Guide (n = 24)

The second interview guide was used for all interviewees of cluster two and three. It was developed when our understanding of data-specific decisions became clearer.

General Background

- What is your educational and professional background?
- What is your role within the organization?

- How does a typical day look like in your routine?
- Where are you in the process of building a data infrastructure for sharing patient data?

Anticipated Technical Solutions

- What data is being collected, processed and shared?
- What is the potential social value of a data infrastructure?
- How can the success of a data infrastructure be measured? (KPI)
- What are incentives to share data with your organization?
- How do ensure data quality?

Organization

- How can a data infrastructure in the making get integrated with your organization?
- What is your organization's data governance structure?
- What kind of consent do you use?
- What is the choice of data sharing architecture within your organization?
- Where is the analysis and processing of the data happening?

#3+ #4 Interview Guide (n = 14)
The third interview guide was used for interviewing participants specifically on ethical values along data related design choices.
General Background

- What is your educational and professional background?
- What is your role within the organization? Since when?
- Where are you in the process of building a health data infrastructure to share patient data inside and outside the organization?

Values

- What do you know about the representation in the data?
- What consent model are you using?
- How do ensure anonymity?
- What are the incentives to improve data quality?

Data-Related Design Decisions & Technical Architecture

- What kind of templates and procedures did you use for your architecture?
- What is the choice of data sharing architecture within your organization?
- Where is the Analysis and Processing of the data happening?
- Who gets access to the data? On what conditions?

Ethical Concern Speed Round:

- Technology leads to equality / Technology leads to inequality
- Prosperity for the individual / Prosperity for society
- More regulation / Less regulation
- Health data belongs to the patient / Health data is common property
- Data sharing should be regulated by the market / Data sharing should be regulated by the state
- With / Without pharmaceutical companies
- Data donation / Money for your data
- Data donation / Obligation to share
- Access to data for all / for a few
- Access to data open / controlled
- Access to data free of charge / access to data for a fee

Appendix 5: Data governance practices – exemplary codes

Symbol coding of primary dimensionality: ƚ technical, ʊ organizational, ∩ normative;
☐ – not present in case; ○ – conceptualized, but not implemented in case; ● – implemented practice in case.

Practice (Purpose & Execution)	Case Example	Selective Codes	
Area of Tension: Data Scope			
Data Scoping (DS) † The practice of segmenting datasets into tailored versions or minimizing datasets to specific metrics resulting in new technical artefacts (new data sets)	**Case Examples:** Starting with two data sets, Alpha scoped data to 28 different data set versions.	**Alpha (●)** **Model driven data selection with feature importance** Alpha, in a network meeting (2024), shows how different data sets are used for different purposes. Including prediction modeling, synthesizing, Explainable AI (XAI) for feature importance. *"There are of course protective measures that are taken, there is of course a form of data minimization, which means that the applicants do not get all the data, but only the data they need for their research project."* - Advisory Board (PO-E) **Gamma** (☐) **Undecided about data selection** *"Should generative AI approaches be allowed and able to use large language models to, for example, naturally linguistic data? Which is certainly available, for example the care entries in the electronic chart, the physician progress reports in the electronic chart, should it also be allowed to use this data or is it only about structured data such as laboratory data, imaging data or other data."* - Participation Expert (Gamma-C)	**Beta (●)** **User-driven data selection with workshops (user = nurses)** Beta, in a network meeting (2024), shows long whiteboards with different risk factors for falls, which have been obtained in user workshops, and were used to narrow the scope of the collected data. Standardized data as well as free text should be made available, since: "*Somehow there is knowledge in institutions about what should be shared.*" - Nurse (Beta-B) **Delta** (○) **Data provider-driven data selection, no own selection implemented** *"This [selection of the use case] was a bit of a lucky coincidence* via *our application partner, you could say, They were able to find a care facility, which we were allowed to visit and which was very open to making something possible… In any case, we decided on the secretion management use case. Secretion management is for ventilated patients who have to be suctioned from time to time if they are ventilated* via *a tracheostomy tube." - Medical IT (Delta-E)* *"These are also different data sources that are supposed to flow together for this: data from the ventilators,* i.e. *the ventilator flow, medication from the patient management system."* - Medical Technology (Delta-B)
Luring with Value (LV) ⊓ **The** practice of offering additional enticement to encourage data providers to share data is informal, addressing norms and motivations of stakeholders	**Case Examples:** Delta used luring value and convinced an additional data provider to share data with the initiative	**Alpha (●)** **Incentives for data provider: cultural shift in organization** "Over the past few years, hospitals have always focused very strongly on the individual patient. This is also how all training was conducted, medical as well as nursing and therapeutic training. The focus is on the person and every patient is individual. and we wanted to **get out of this perspective** and say perhaps there are larger **patterns** that we can recognize if we process the data better, if we simply learn to handle our data better. and that was the main motivation to take part in the project." - CEO Clinic II (Alpha-E) University Clinic 1 and Geriatric Clinic 2 generated technical artefacts such as learnings on data collection practices, which were documented for future projects and quality management (Workshop, February 2023). **Gamma** (☐) **No incentives** Even after a year, Gamma did not	**Beta** (○) **Incentives for data provider: Competitive advantage** *"we (data provider) expect **process optimization** in any case. Process optimization also creates **added value,** because processes actually run more smoothly, the quality of care improves and there is a **competitive advantage**… to involve the quality management of the department to see what workflows can perhaps be developed how can we look again at where we perhaps need to do more training, …knowledge management and yes process improvement through perhaps adapted workflows"* **-** Nurse (Beta-B) At the end of the observation period, the process had not resulted in tangible outcomes. Therefore, we coded this as (○). **Delta (●)** **Incentives for data subject: additional information on**

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		receive the promised data from the clinic's data hub due to lack of incentives for the data provider: “*so it [data providing center] was still very research-driven, and the focus of the data that was retrieved was of course also closely linked to these scientific projects, but now it merged to a data migration program at our central institution, so the center has no longer any direct incentives, so to speak*.” - Medical Doctor and Medical IT expert (Gamma-A) Gamma further evaluated potential values of the data with a user survey: *“The survey with the nursing staff and, above all, also reflected by the interviews with the doctors, is that there is a* ***primarily interprofessional communication problem in the treatment of delirium,*** *because there are simply very different views on the point at which treatment is required and which treatment standards should actually be applied in early detection prevention.”* - Participation Expert (Gamma-C)	**research outcomes** Delta considered data donation as a model, while one established initiative led as a data sharing enabler: *“we received a lot of letters to our e-mail address from people who asked questions about the project somehow, whether you can donate more data, because Covid is something, especially with long Covid, people have the feeling that I'm not being seen properly here by my doctors and now maybe this project can bring me something for my personal outcome”* - Medical IT expert (DSE—D) **Incentives for data providers: new insights from market data** *“when you get data from others that is valuable to you, so that would be the incentive for us, and that you would have the AI model trained locally on your data, for example. That others would train their AI models on our data* via *federated learning.”* - Medical Technology (Delta-B) **Incentives: tangible value added** Delta concluded: especially in the rare disease area, patients are ´desperate´ for their data to be used in a meaningful way to develop new therapies, – workshop 2023 “*So incentives, I always have a bit of a problem with that, because data is valuable, I know that, but I wouldn't want to go into that level at all, that would be something ethically again, something critical for me would be exactly what I had just argued, this work of persuasion, when we collect data,* ***we collect this data to change the system. … It also has to be tangible.*** *I then have to see, aha, it really had a benefit that I got involved and gave away data.*” - Nursing Science (Delta-D) In a results report at the end of the project (2025), Delta stated that they “convinced” a previously unrelated private company to integrate their data. - project website, June 2025
Evaluating Algorithmic Fairness (EF) [‡] The practice of implementing empirical evaluation frameworks to assess data and AI-driven outcomes to address biases in data and models results in new technical artefacts (dashboards, metrics etc.)	**Case Examples:** Empirical evaluation of fairness measures in synthetic data in Alpha	**Alpha (●)** **Evaluation means training on all data, evaluating outcome for specific groups** *“in the context of health data and our project in particular, some factors of data quality that influence the prediction quality are especially important: it could be that certain population groups are underrepresented in the data or that there are some shifts in the data that are associated with belonging to an underrepresented group. This kind of data quality problem can then lead to poorer predictions by the machine, but only if you belong to a certain group.*” - Machine Learning expert (Alpha-C) Alpha represented Explainable AI methods in presentations as a possible	**Beta** (☐) Was not relevant in this case.

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		technological practice to indicate relevant fairness metrics. *We also look at the extent to which the predictions of our fall risk models. For these respective groups.* ***So fair is representative, and fair is not, so to speak, interchangeable.*** *But in order to see whether the data situation is representative enough to train a model that then works equally well on all groups, for example old women or younger people, for example men, in order to check this,* ***we first train a model on all data or on certain data sets and then see how well the model that we have trained on this data works.*** (Alpha-C)	
		Gamma (○) **Evaluation means training on all data, evaluating outcome for noise** Interviewer: Do you have quality monitoring in terms of the representativeness of the data? *“a very good starting point for quality measures is simply to* ***plot the data over time and see. Are there any fluctuations or does it look relatively random*** *or is it perhaps a bit seasonal, but the basic effect is* ***noise*** *or do I suddenly have any ugly spikes or gaps in the data? You can see that relatively quickly”* - Medical Doctor and Medical IT expert (Gamma-A)	**Delta** (○) **Evaluation means anomaly detection** *“Anomaly detection means we get data streams and it becomes clear that there are actually only male data donors, and that is of course bad for drug development or something like that and then we could give that as an indication. “Watch out! the representativeness is not given here.”* - Delta-E Other means for evaluating fairness have not been detected.
Contextualizing Data Quality (CQ) [‡] The practice of providing contextual indicators that help stakeholders assess data quality without altering the original dataset results in new technical artefacts (e.g., protocols)	**Case Examples:** Automated anomaly detection in Delta	**Alpha** (○) **for evaluating similarity of synthetic data** Alpha commented on data quality in regards to their attempts to synthesize data. Data quality was linked to privacy. *“my task is to outline this technical solution, .., with a focus on synthetic data and, above all, the extent to which data protection requirements can be made empirically verifiable with methods that we are developing.”* - Machine Learning expert (Alpha-C) Later in the process, Alpha presented their first outcomes (workshop notes, Oct 2024): “Synthetic data should be: 1. similar to real data 2. privacy preserving 3. useful for training ML models. Results comparing six data sets: Utility of synthesized data depend heavily on the original data”	**Beta** (☐) Was not highlighted in this case, however *[Vendor X] already knew that the data is “bad” (Barcamp notes, session on Data Governance)*
		Gamma (●) **for reinforcement learning and clinical usability** Gamma considered clinical feedback as an essential part of reinforcement learning, one predictor of good data quality: *“The results of the prediction assessment also go back in, you can also learn from it again if the algorithm repeatedly receives feedback from the clinicians ‘had no delirium’, this is totally important information to be able to train”* - Medical Doctor and Medical IT expert	**Delta (●)** **for evaluating data on AI usability** Delta considered only unprocessed data, using a machine-learning focused approach: *“We focus on anomaly detection, … we try to recognize whether something is unusual in the data and would add this to the data set as a comment, so to speak.* ***We don't change the data,*** *because we don't know what the people who download it want to do with it, but we do want to give an indication”* – Project manager (Delta-A)

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		(Gamma-A) Gamma contextualized data quality with feedback from clinical experts: *"It is very important for the quality of the data, when you open it up, that you look into the data and evaluate it to find inconsistencies or special features and then compare them immediately. Are the characteristics of the data also in line with reality or do I perhaps have an artifact and this is best done in a specialist project, where we now also have the nursing experts in the group"* - Medical Doctor and Medical IT expert (Gamma-A)	*"..we want to give a statement about the data quality in any form to the person who downloads the data: They.. can then decide for themselves whether the data is still sufficient."* - Medical IT (Delta-E)
Synthesizing Data (SD) ⸸ The practice of transforming original data into synthetic data to balance data privacy and usability results in <u>new technical artefacts.</u>	**Case Examples:** Alpha used synthetic data for mitigating issues with anonymity, and used fairness metrics for mitigating issues with bias	**Alpha (●)** **Synthetic data as representative / similar data to original** In the project plan of Alpha, synthetization was considered one of the key mechanisms to ensure data privacy and anonymity of the patients. (project plan). The advisory board member explained the advantage: *", ..., so that* ***you practically represent a world*** **in silico** *like in a flight simulator. ...and it makes no difference whether you take the purely instrumental unprocessed data or such complexly constructed digital avatars."* – Governance expert (PO-A) The data science lead added: *"We train models that can generate synthetic data. Tabular data that are similar enough to real data. They are similar enough that they can be used to train meaningful machine learning models, ideally those models work just as well on real data as machine learning models trained on real data. at the same time these synthetic data should not reveal personal data or sensitive information about patients."* - Machine Learning expert (Alpha-C)	**Beta** (☐) Was not relevant in this case.
		Gamma (☐) Was not relevant in this case.	**Delta** (○) ***Synthetic data as condition for data sharing with competitors*** *Delta aimed to use synthetization to get access to data by data providers and use Federated Learning as one privacy enhancing practice:* *"The data provided directly by the vendor is not real data, but simulated data that they use anyway with simulation tools as a medical technology manufacturer."* - Medical IT (Delta-E) "*The other option is to keep the data with you, but release it for federated learning, so to speak.*" - Medical Technology (Delta-B)
Exercising Research Privilege (RP) ⊓ The practice of leveraging Section 9, GDPR provisions to enable access to health data for research purposes in the public interest without explicit consent is a	**Case Examples:** Alpha used the research privilege as an argument to share synthetic data, Alpha used synthetic data as an argument to share data without explicit consent obtained	**Alpha (●)** **Combinatory conditions to rely on research privilege (RP)** **From Clinic 1 and Clinic 2, routine data was used where patients did not give explicit consent. As one data sharing enabler in exchange with Alpha explains:**	**Beta (●)** **Non-for-profit purpose to rely on RP** In a barcamp in 2024, Beta reported that it switched from a consent model, which was originally planned, to **exercising the research privilege**. (Barcamp notes, session

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
direct response to the regulatory landscape.		*"The entire architecture that we have agreed with the supervisory authority is based on the fact that* ***we are not aiming for the individual consent of the patient, but that we do many other things that, in combination,*** *ultimately allow us to do this without consent."* CEO Data Sharing Company - DSE-A An advisory board member explained further, how the GDPR regulation states data processing without consent: *"Ultimately, broad consent is not necessary in principle, you could also give specific consent and then use article six or article nine (of the GDPR) and carry out this* ***repurposing****, so on this legal basis you could then also use the data secondarily for other research purposes, …"* – Governance expert (PO-A)	Data Governance) *"One thing that has changed. We now want to use the data* ***on the basis of the research privilege, which means that we don't have to obtain consent****, but we can justify that this is non-profit research."* - Nursing Science (Delta-D) **Additional information for data subjects to rely on RP** *"…of course this is very sensitive data and we will very probably invoke the research privilege on the one hand and on the other hand we will then inform our residents that we are still using their relevant information from the documentation system for a project.*" - Nursing Science (Delta-D)
		Gamma (○) **Non-for-profit condition to rely on RP** **Not using consent allows the projects to get access to retrospective data analysis:** *"Basically, the patient is not part of the study, it is a retrospective data collection, at least if you are training the AI algorithm …* ***I don't ask the patient for advice if I think it's medically correct to do so****"* - Medical Doctor and Medical IT expert (Gamma-A)	**Delta** (☐) **No RP** *In a barcamp in 2024, Delta reported that it used* ***earmarked (specific) consent,*** *meaning each patient agreed individually to share their data. Leading to the inclusion of 12 patients in their study.* (Barcamp notes, session Data Governance)
Leveraging Parents (LP) $^{\bar{\circ}}$ The practice of collaborating with or utilizing the resources, expertise, or influence of large, established organizations, thereby using relational mechanisms to build trust in protective data sharing and procedural mechanisms to gain efficiency in negotiations (e. g. scalable processes).	**Case Examples:** Beta and Gamma aimed to adopt the policies of their parent organization, for instance its consent model or secure storage concept	**Alpha** (○) **One of the project organizations as secure infrastructure hosts** The Interviewer asked an advisory board member what leads to building trust: *"To have seals of approval …and the name of the clinic always has a good trust bonus"* - Alpha-F "A*s long as we provide the data on the university clinic's infrastructure (*Clinic 1)*, I can also get the Board* (of Clinic 2) *to agree to this."* - Clinic 2 CEO, Alpha-E (Notes to a Board of Directory meeting in July 2024) **Open science infrastructure providers as hosts for the data** One advisor from the university hospital advised that the synthetic data was strongly anonymized, risks are negligible, and that data can be securely uploaded to a public infrastructure such as Zenodo	**Beta** (●) **Parent organizations as secure infrastructure hosts** In a barcamp, Beta presents an architectural chart where most of the existing infrastructure from the university clinic is used. There is a red database for the existing FHIR database of the clinic and a newly added blue one for the project, showing the "distinction" between the current project and existing research infrastructure. **Parent organizations as means of efficiency and scalability** *"There is no longer a lot of negotiating of contracts, instead a contract template is used that has already been agreed in advance with all locations. The application is simply attached and then signed and the project can get underway."* - Medical IT expert at consortia (PO-D)
		Gamma (●) **Parent organizations as means of efficiency and scalability** *"I think it is imperative that we do not create completely new concepts for such practical data projects*." - Medical Doctor and Medical IT expert (Gamma-A) **Parent organizations as secure infrastructure hosts** *"In any case, the benefit for the data integration center and thus the location is that we have the data already described in the core dataset available in real time.*	**Delta** (☐) Was not relevant in this case.

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		The second step would be for us to try to bring the parameters not described in the core dataset into the discussion as a corresponding FHIR profile or at least to link them to the consortia structures to see how we can create a core dataset for delirium or geriatrics similar to the Covid dataset, which is also profiled, which is also a good result, that we say these are the delirium-relevant parameters here. However, balancing something like this can be very time-consuming, It's more ***about docking the project to the existing structures of the consortia*** *and not trying to recreate what the consortia does on a small scale"* - Medical Doctor and Medical IT expert (Gamma-A) **Responsibility stays with each data sharing initiative** *"I strongly believe that we should take responsibility for our own actions. If we decide that data is relevant, we must justify this decision well, make the potential risks transparent, and explain how we will deal with them."* - Nursing Scientist (Gamma-B)	
Broadening Consent (BC) [∩] The practice of obtaining broad consent from patients that encompasses broad research purposes is a **direct response to the regulatory landscape**.	**Case Examples:** Gamma used the broad consent to share pseudonymized (not only anonymized) data, yet had to incentivize data subjects to agree	**Alpha (●)** Clinic 2 broadened its consent to allow for AI use cases, which was not possible under the old consent regime. *"We learned a lot during the project and have revised our patient treatment contract and our general terms and conditions with regard to broad consent."* - Clinic 2 CEO (Alpha-E)	**Beta** (○) Was considered but not done. *"We came across a French guideline that describes in detail how to train AI systems in a consent-based manner that complies with the GDPR. It recommends obtaining two separate consents to train an AI system: one for the purpose of advancing research, and the other for the purpose of improving care. We considered applying this approach.."* - Medical IT expert (Beta-A) *"The broad consent is strongly focused on research and is not entirely comprehensible to patients and nursing staff."* - Nursing Scientist (Beta-B)
		Gamma (●) The university clinic revised its consent policy to include AI development into its consent policy.	**Delta** (☐) Not envisioned.
Separating Storage (SS) [ǂ] The practice of physically separating data results in new technical artefacts (i.e. new, separate servers)	**Case Examples:** Beta: two separate environments for anonymized and pseudonymized data	**Alpha (●)** **Original data remains in separate local infrastructure, synthetic data is shared** Alpha installed a separate computer on-site in the geriatric hospital that was not connected to the clinic network in order to secure the server from cyber attacks. *"We don't want datasets to leave our house and I think that's what we would still aim for. So, we wouldn't like to let data leave the house. You never know what happens to it."* - (Alpha-E) However, an advisory board member argued against this approach: *"A large IT provider can provide a more secure IT infrastructure than, for example, a university hospital with its limited resources. In this respect, I would say that the future clearly lies in the cloud."* - (DSE-E) The Clinic was hacked during the	**Beta (●)** **Original data remains in separate local infrastructure, models are shared** The main consortia provided the option to not only share data with potential users but share models with potential data providers, without data leaving the facility (distributed analyses): "*We also have to consider how they want to analyze the data, whether they need the data for the analysis on their own computer or whether they might provide us with a script with which the data can be analyzed on site..*" - Medical IT (PO-D) **Storage structure is linked to consent** "*The main argument is always consent*" - Medical IT (Gamma-H) meaning that **only data with consent can be shared.** *"It is always an advantage to*

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		project duration, there were no indications that data was stolen (internal protocols) **Gamma (●)** **Original data remains in local infrastructure, AI training is installed on premise** Gamma aimed to integrate different heterogenous dataset with interoperable standards such as FHIR. Gamma further decided to not separate storage for secondary data usage, but **install a FHIR store within the primary network:** *"We want to evaluate the data as soon as it is generated, and that's how the idea came about, also inspired a little by companies that offer AI systems on routine data: we establish our own FHIR server in the hospital information system itself. To be able to make such decision support systems in real time. Ultimately these are systems that are part of patient care and therefore also belong in the hospital information systems solely in terms of support and maintenance, also solely in terms of responsibilities. ...A ready-made algorithm would then run on this server, which in turn would display the result of the prediction to authorized persons."* - Medical Doctor and Medical IT expert (Gamma-A)	*do distributed evaluation because you can access a larger amount of data because you don't have consent for all the data."* - Medical IT (Gamma-H) **Delta** (○) **Undecided on storage solutions** *"It won't be one platform, but different systems that network with each other."* - (Delta-E) *Delta also considered Federated Learning: "A very exciting aspect that one consortium partner is currently pursuing is federated learning, where it is possible to pre-select and pre-process data in a decentralized manner, and that then perhaps only something that is less critical from a security point of view is stored centrally-"* - Medical IT (Delta-E)
Opening Knowledge (OK) ⊓ The practice of openly sharing aggregated data results while withholding actual individual-level patient data is informal, addressing norms and motivations of stakeholders	**Case Examples:** Alpha used opening knowledge as a practice to mitigate timing issues until actual data could be shared, Delta used it as a condition for data users to gain data access	**Alpha (●)** **Open Science as principle** "*The opportunity for researchers to work with such data was one of the main motivations for me and hopefully this will have a **catalyzing effect** if we can make it possible to make data shareable through the technical development we are advancing.*" - Machine Learning expert (Alpha-C) Alpha repeated their dedication to open science and open knowledge by writing articles on Wikipedia and working on blueprints and templates for anonymization proceedings (Workshop, 13.02.2023) Another role model from an established data sharing platform explained: **Open knowledge as enabler of trust** "*we tried to use our science blogs to explain in as small a way as possible what we are doing, why we are doing it there were always questionnaires, for example, that didn't just appear with the app, but there was a PR campaign where it was shown that there are now the following new functions that we are bringing in, you don't have to use them…that was an important learning: **the communication about what we are doing with the data, why we are doing this with the data.***" - Medical IT expert (DSE-D) **Gamma** (○) **Open knowledge as principle (for social innovation)** Gamma defined the provision of open	**Beta** (☐) **Open knowledge as enabler of best practice** "*We are still discussing daily how we should go about obtaining consent for data processing because with AI, we are still on relatively uncharted territory.... we have come across a guideline in France that describes relatively step by step how to design AI applications or the use of AI applications in accordance with consent.*" - project manager (Beta-A) The French guideline reads as follows: "*On 5 April 2022, the French Data Protection Authority ('CNIL') released extensive publications concerning artificial intelligence (AI)0.1 These publications are addressed to three main target groups, namely the broader public, experts, and scientists, as well as organizations that process personal data through AI systems.*" (DataGuidance, 2022) **Delta** (○) **Open knowledge as enabler of best practice** "*It would be exciting if I could also use*

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		knowledge and its innovation not solely on a technological level: "*There is a certain pressure to innovate that prevails and is already evident at the level of funding logic, where one would have to say, yes, one should also be able to promote projects that somehow promote social innovations or develop social solutions. And that innovation should not always be seen in the narrow sense of technology, but that process-oriented or social interventions can be just as innovative. Interventions can be just as innovative as technology.*" - Participation Expert (Gamma-C)	*the platform when the AI developer has developed his solution with it, that I can perhaps make my AI model available again as a starting point for further development by another team, perhaps I have the opportunity to offer a license on how to continue using my model*" - Delta-E **Open knowledge as enabler of user value on data sharing infrastructure** *"No fee (to receive data) is paid, but the payment is the knowledge that was generated from this data, to it available to the others or to the platform."* - Delta-E
Engaging Communities (EC) Ꝺ The practice of integrating stakeholders in decision-making processes about the use of data, thereby using relational mechanisms to build trust	**Case Examples:** Beta consulted with users in committees to enhance AI potential, Gamma used a Use and Access committee	**Alpha** (□) **Participatory concepts as cultural shift** The expert reported from a best practice example that was used as a role model: *"A central competence center for shared decision making was appointed in order to really have empowered patients.* i.e. *participatory decision-making effects are on everyone's lips, but on the other hand is there of course also a certain resistance to new concepts in the healthcare system."* - Medical Technology expert (Alpha-A) However, did not play a main role. **Gamma (●)** **Participatory concepts for problem identification** Gamma considered inputs from nursing staff and doctors: *"we try to discuss ethical problems, then of course that is also taken into account. Yes, I would say with caution in the discussions, but in my view this has little influence or little to do with the formal process of whether or not data may be used."* – Participation scientist (Gamma-C) A bigger issue was highlighted: "*The technology is already taken for granted and is not looked at, at all: what is actually the complex social problem that has to be improved through the use of technology, i.e…It often turns out, as in this project, that a large proportion of the problems identified … They do not need to be solved technically, because they are* ***primarily interactional social or communication problems that could or perhaps even should be tackled without technology.*** "- Participation scientist (Gamma-C)	**Beta (●)** **Participatory concepts to gain domain knowledge** Nurses were interviewed on data documentation, processes and usage in order to highlight relevant use cases and prolong with a minimized dataset for sharing. "*which risk factors are considered important by nursing staff and are they documented digitally or are they not documented digitally, is it necessary to document them digitally or are things documented that are not useful and necessary?.*" - Nursing Scientist (Beta-B) **Delta** (○) **Participatory concepts for user motivation** A nursing expert explained why keeping end-users and other important stakeholders engaged in the evaluation process is important: *"Nursing staff are sometimes highly motivated, and are really extremely positive about new technologies, but then find it difficult to enter everything by hand somewhere, or it is not accepted. And then, of course, the tolerance for such new things becomes less and less in order to introduce something new."* - Nursing Scientist (Delta-C)
Layering Access Control (LC) Ꝺ The practice of layering access to data along affiliation criteria, thereby using procedural mechanisms to obtain data release	**Case Examples:** Depending on consent, different access rights at Beta and Gamma, for Delta, OK was a condition to gaining data access	**Alpha** (○) **Access process highlighting affiliations** An advisory board member shares their best practice example how access to data is a layered process: *"You must submit an application… We collect an incredible number of variables and the users can't just say we're going to take everything on the subject of falls. Each case has to be justified*	**Beta (●)** **Access process of parent organization, highlighting affiliations** *"And in order to guarantee data protection, it is of course the case that the locations, for example, have* ***trustee offices,*** *local trustee offices, which then also know which patient is behind each patient identifier."* - Medical IT (PO-D)

(*continued on next page*)

(*continued*)

Practice (Purpose & Execution)	Case Example	Selective Codes	
		individually." - (Alpha-F) Later, the expert emphasized the prioritization of affiliated research requests: *"Internal, affiliated data must also be requested on a central platform. So of course that will be prioritized. We get internal data usage notifications while external data usage applications must be submitted."* - (Alpha-F) The data scientist expert adds: "*Typically, you have to register* via *an e-mail address that is assigned to a research institution.*" - (Alpha-C) However, this process requires intensive resources and receives also critical feedback: *"We often get complaints that the application procedure is too complex,* i.e. *that they have to provide too much information."* - (Alpha-F)	"*When registering, they have to provide us with some kind of reference. For most users, this is simply the affiliation to a research institution, but international researchers are also likely to register or from companies – we are not limited to public research*" - (PO-D) **Technical layer of the access process: feasibility requests** *"However, we have a so-called feasibility request. This is a technical possibility for users to ask roughly, what do you have, do you have what I need, where you can specify your inclusion and exclusion criteria for a project request and receive a fully automated response."* - (PO-D)
		Gamma (●) **Access process of parent organization, highlighting affiliations.** **Organizational layer of the access process: expert committees** *"We have to adhere to the processes of the data integration center and data use applications that are currently being established, …Our process: you don't get ethics without a data protection vote, we would first talk to the data protection officer, if they wave it through, it goes to ethics. Once we've gone through ethics, we have a committee, these are all structures that the consortia has planned"* - (Gamma-A)	**Delta** (○) **Technical layer of the access process: adding value with technical artifact** *"the first question that arises is: there a controlled access that someone has to provide us with something beforehand ..: do you then get full access to whatever you want from the platform or do you first get, for example, a test data set or a small data set and* ***have to prove yourself,*** *so to speak, by making the trained model available to the platform again. No fee is paid, but the* ***payment is the knowledge*** *that was generated from this data, to it available to the others or to the platform."* - (Delta-E)

Data availability

Data will be made available on request.

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The making of data commodities: Data analytics as an embedded process. *Journal of Management Information Systems, 38*(2), 401–429. https://doi.org/10.1080/07421222.2021.1912928

Aaltonen, A., Alaimo, C., Parmiggiani, E., Stelmaszak, M., Järvenpää, S. L., Kallinikos, J., & Monteiro, E. (2023). What is missing from research on data in information systems? Insights from the inaugural workshop on data research. *Communications of the Association for Information Systems, 53*, 475–490. https://doi.org/10.17705/1CAIS.05320

Aaltonen, A., & Stelmaszak, M. (2023). The performative production of trace data in knowledge work. *Information Systems Research, 35*(3), 1448–1462. https://doi.org/10.1287/isre.2019.0357

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. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Alaimo, C., & Kallinikos, J. (2024). Decentering organizations: Data, knowledge, and institutional change. In *, Chap. 6. Data rules: Reinventing the market economy*. MIT Press. https://doi.org/10.7551/mitpress/11751.003.0011.

Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023). Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization. *Information and Organization, 33*(3), Article 100478. https://doi.org/10.1016/j.infoandorg.2023.100478

Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly, 31*(1), 78–108. https://doi.org/10.2307/2392767

Bauer, K., Heigl, R., Hinz, O., & Kosfeld, M. (2024). Feedback loops in machine learning: A study on the interplay of continuous updating and human discrimination. *Journal of the Association for Information Systems, 25*(4), 804–866. https://doi.org/10.17705/1jais.00853

Benfeldt Nielsen, O. (2017). A comprehensive review of data governance literature. *IRIS: Selected Papers of the Information Systems Research Seminar in Scandinavia, 8*(3), 120–133. https://aisel.aisnet.org/iris2017/3.

Benfeldt, O., & Persson, J. S. (2025). Semiotic mediation in data governance: Towards valuing data as assets. *Information and Organization, 35*(3), Article 100588. https://doi.org/10.1016/j.infoandorg.2025.100588

Black, S., Davern, M., Maynard, S. B., & Nasser, H. (2023). Data governance and the secondary use of data: The board influence. *Information and Organization, 33*(2), Article 100447. https://doi.org/10.1016/j.infoandorg.2023.100447

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: A systematic review of data governance principles. In H. J. Scholl, et al. (Eds.)*, Vol. 9820*. *Electronic government* (pp. 115–125). Springer International Publishing. https://doi.org/10.1007/978-3-319-44421-5_9.

Burmann, A., Langer, H., Bosompen, J., Clemens, T., Herten, B., Kerstan, L., Lauer, R., Naeve, P., Geller, S., Huhnt, J., & Merkel, S. (2023). Governance of digital health data on cooperatively organized platforms – A design thinking approach. In *proceedings of the 56th Hawaii international conference on system sciences*. IEEE Computer Society. https://doi.org/10.24251/HICSS.2023.757.

Cao, S., & Iansiti, M. (2021). Data governance, interoperability and standardization: Organizational adaptation to privacy regulation. *Harvard Business School Working Paper, 21*(122).

Celiktutan, B., Klesse, A.-K., & Tuk, M. A. (2024). Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations. *International Journal of Research in Marketing, 41*(3), 496–512. https://doi.org/10.1016/j.ijresmar.2024.05.006

Cepa, K., & Schildt, H. (2023). Data-induced rationality and unitary spaces in interfirm collaboration. *Organization Science, 34*(1), 129–155. https://doi.org/10.1287/orsc.2021.1566

Chen, X., Li, G., Li, S., & Zheng, Q. (2024). When should a sharing platform adopt the bilateral review system? *MIS Quarterly, 48*(3), 1121–1156. https://doi.org/10.25300/MISQ/2023/17596

Costabile, C., & Øvrelid, E. (2023). Identifying governance mechanisms for data sharing in collaborative platform ecosystems. *Proceedings of the European Conference on Information Systems (ECIS), 2023*(133). https://aisel.aisnet.org/ecis2023_rp/283/.

DataGuidance. (2022). France: CNIL guidance on AI - A preview of the EU AI Act?. https://www.dataguidance.com/opinion/france-cnil-guidance-ai-preview-eu-ai-act.

Davidson, E., Wessel, L., Winter, J. S., & Winter, S. (2023). Future directions for scholarship on data governance, digital innovation, and grand challenges. *Information and Organization, 33*(1), Article 100454. https://doi.org/10.1016/j.infoandorg.2023.100454

Dyke, S. O. M., Linden, M., Lappalainen, I., de Argila, J. R., Carey, K., Lloyd, D., … Flicek, P. (2018). Registered access: Authorizing data access. *European Journal of Human Genetics, 26*(12), 1721–1731. https://doi.org/10.1038/s41431-018-0219-y

European Commission. (2021). *European Commission white paper on artificial intelligence–a European approach to excellence and trust*.

Fayard, A.-L., Gkeredakis, E., & Levina, N. (2016). Framing innovation opportunities while staying committed to an organizational epistemic stance. *Information Systems Research, 27*(2), 302–323. https://doi.org/10.1287/isre.2016.0623

Foster, J., McLeod, J., Nolin, J., & Greifeneder, E. (2018). Data work in context: Value, risks, and governance. *Journal of the Association for Information Science and Technology, 69*(12), 1414–1427. https://doi.org/10.1002/asi.24105

Friedrich, T., Akbari, K., & Fuerstenau, D. (2026). Data governance practices for generative AI powered organizational knowledge management systems using retrieval-augmented generation. In *Proceedings of the 59th Hawaii international conference on system sciences (HICSS)*.

Frühwirth, M., Pammer-Schindler, V., & Thalmann, S. (2024). Knowledge leaks in data-driven business models? Exploring different types of knowledge risks and protection measures. *Schmalenbach Journal of Business Research*. https://doi.org/10.1007/s41471-024-00189-z. Advance online publication.

Gerring, J. (2007). *Case study research: Principles and practices*. Cambridge University Press. www.cambridge.org/9780521859288.

von Grafenstein, M. (2022). Reconciling conflicting interests in data through data governance. An analytical framework (and a brief discussion of the data governance act draft, the data act draft, the AI regulation draft, as well as the GDPR). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4104502

Gray, B. L., Briscoe, F., & Diaz Ferraro, C. (2022). The technological entrainment of moral issues: The case of genomic data markets. *Academy of Management Journal, 66*(4), 1123–1151. https://doi.org/10.5465/amj.2019.1202

Grodal, S., Anteby, M., & Holm, A. L. (2021). Achieving rigor in qualitative analysis: The role of active categorization in theory building. *Academy of Management Review, 46*(3), 591–612. https://doi.org/10.5465/amr.2018.0482

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Hernández, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2022). Synthetic data generation for tabular health records: A systematic review. *Neurocomputing, 493*(7), 28–45. https://doi.org/10.1016/j.neucom.2022.04.053

Hillebrand, K., Hornuf, L., Müller, B., & Vrankar, D. (2023). The social dilemma of big data: Donating personal data to promote social welfare. *Information and Organization, 33*(1), Article 100452. https://doi.org/10.1016/j.infoandorg.2023.100452

Hofmann, J., Katzenbach, C., & Gollatz, K. (2017). Between coordination and regulation: Finding the governance in internet governance. *New Media & Society, 19*(9), 1406–1423. https://doi.org/10.1177/1461444816639975

Jakubik, J., Vössing, M., Kühl, N., Walk, J., & Satzger, G. (2024). Data-centric artificial intelligence. *Business & Information Systems Engineering, 66*(4), 507–515. https://doi.org/10.1007/s12599-024-00857-8

Jarke, M., & Quix, C. (2022). Federated data integration in data spaces. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing data spaces: The ecosystem approach to competitive advantage* (pp. 181–194). Springer. https://doi.org/10.1007/978-3-030-93975-5_11.

Jarvenpaa, S., & Essén, A. (2023). Data sustainability: Data governance in data infrastructures across technological and human generations. *Information and Organization, 33*(1), Article 100449. https://doi.org/10.1016/j.infoandorg.2023.100449

Jung, C., & Dörr, J. (2022). Data usage control. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing data spaces: The ecosystem approach to competitive advantage* (pp. 129–146). Springer. https://doi.org/10.1007/978-3-030-93975-5_8.

Jussen, I., Möller, F., Schweihoff, J., Gieß, A., Giussani, G., & Otto, B. (2024). Issues in inter-organizational data sharing: Findings from practice and research challenges. *Data & Knowledge Engineering, 150*, 1–19. https://doi.org/10.1016/j.datak.2024.102280

Kari, A., Bellin, P., Matzner, M., & Gersch, M. (2025). Governing the emergence of network-driven platform ecosystems. *Electronic Markets, 35*(1), 4. https://doi.org/10.1007/s12525-024-00745-9

Kearns, M., & Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152. https://doi.org/10.1145/1629175.1629210

Khurshid, A., Rousseau, J. F., Andrews, S. B., & Tierney, W. M. (2020). Establishing a data-sharing environment for a 21st-century academic health center. *ACI Open, 04*(01), e59–e68. https://doi.org/10.1055/s-0040-1709652

Kilgus, T., Patecka, A., Schurig, T., Kari, A., Gubser, R., Gersch, M., … Fürstenau, D. (2024). Creating value from the secondary use of health data: International examples, best practices, and opportunities to scale. *Communications of the Association for Information Systems, 55*, 507–534. https://doi.org/10.17705/1CAIS.05520

Klievink, B., van der Voort, H., & Veeneman, W. (2018). Creating value through data collaboratives: Balancing innovation and control. *Information Polity, 23*(4), 379–397. https://doi.org/10.3233/IP-180070

Lefebvre, H., Krasikov, P., Legner, C., & Flourac, G. (2025). Data management as a joint value proposition—A design theory for horizontal data sharing communities. *Electronic Markets, 35*(1). https://doi.org/10.1007/s12525-025-00755-1

Lis, D., & Otto, B. (2020). Data governance in data ecosystems – Insights from organizations. In *Proceedings of the 26th Americas conference on information systems (AMCIS 2020)*.

Lukyanenko, R., Parsons, J., Wiersma, Y. F., & Maddah, M. (2019). Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Quarterly, 43*(2), 623–647. https://doi.org/10.25300/MISQ/2019/14439

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems, 19*(12), 1253–1273. https://doi.org/10.17705/1jais.00526

Mayring, P., & Fenzl, T. (2014). Qualitative inhaltsanalyse. In N. Baur, & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 543–556). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18939-0_38.

Meuser, M., & Nagel, U. (2009). The expert interview and changes in knowledge production. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 17–42). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_2.

Mikalsen, M., & Monteiro, E. (2021). Acting with inherently uncertain data: Practices of data-centric knowing. *Journal of the Association for Information Systems, 22*(6), 1715–1735. https://doi.org/10.17705/1jais.00722

Mühlhoff, R., & Ruschemeier, H. (2024). Updating purpose limitation for AI: A normative approach from law and philosophy. *International Journal of Law and Information Technology, 33*. https://doi.org/10.1093/ijlit/eaaf003

Noesgaard, M. S., Nielsen, J. A., Blegind Jensen, T., & Mathiassen, L. (2023). Same but different: Variations in reactions to digital transformation within an organizational field. *Journal of the Association for Information Systems, 24*(1), 12–34. https://doi.org/10.17705/1jais.00770

Otto, B., ten Hompel, M., & Wrobel, S. (Eds.). (2022). *Designing data spaces: The ecosystem approach to competitive advantage*. Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5.

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems, 34*(2), Article 101885. https://doi.org/10.1016/j.jsis.2024.101885

Paparova, D., Aanestad, M., Vassilakopoulou, P., & Bahus, M. K. (2023). Data governance spaces: The case of a national digital service for personal health data. *Information and Organization, 33*(1), Article 100451. https://doi.org/10.1016/j.infoandorg.2023.100451

Parmiggiani, E., & Grisot, M. (2020). Data curation as governance practice. *Scandinavian Journal of Information Systems, 32*(1), 3–38. https://aisel.aisnet.org/sjis/vol32/iss1/1.

Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2022). In the backrooms of data science. *Journal of the Association for Information Systems, 23*(1), 139–164. https://doi.org/10.17705/1jais.00718

Parsons, J., Lukyanenko, R., Greenwood, B., & Cooper, C. (2025). Understanding and improving data repurposing. *MIS Quarterly*, 1–35. https://doi.org/10.25300/MISQ/2025/18361

Patecka, A., Parmiggiani, E., & Wessel, L. (2025). Context matters? Contextual data quality for data repurposing. In *Proceedings of the European conference on information systems (ECIS) 2025, Amman, Jordan*.

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in the age of algorithms. *Health Technology, 7*, 351–367. https://doi.org/10.1007/s12553-017-0179-1

Prinz, W., Rose, T., & Urbach, N. (2022). Blockchain technology and international data spaces. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing data spaces: The ecosystem approach to competitive advantage* (pp. 165–180). Springer. https://doi.org/10.1007/978-3-030-93975-5_10.

Pujol Priego, L., Wareham, J., & Romasanta, A. K. S. (2022). The puzzle of sharing scientific data. *Industry and Innovation, 29*(2), 219–250. https://doi.org/10.1080/13662716.2022.2033178

Putnam, L. L., Fairhurst, G. T., & Banghart, S. G. (2016). Contradictions, dialectics, and paradoxes in organizations: A constitutive approach. *The Academy of Management Annals, 10*(1), 65–171. https://doi.org/10.1080/19416520.2016.1162421

Rouse, E., Reinecke, J., Ravasi, D., Langley, A., Grimes, M., & Gruber, M. (2025). Making a theoretical contribution with qualitative research. *Academy of Management Journal, 68*(2), 257–266. https://doi.org/10.5465/amj.2025.4002

Sabatello, M., Martschenko, D. O., Cho, M. K., & Brothers, K. B. (2022). Data sharing and community-engaged research. *Science, 378*(6616), 141–143. https://doi.org/10.1126/science.abq6851

Safadi, H., Skousen, T., & Karahanna, E. (2024). Firm-sponsored online communities: Building alignment capabilities for participatory governance. *Information Systems Research, 36*(1), 419–439. https://doi.org/10.1287/isre.2021.0578

Sandberg, J., & Alvesson, M. (2020). Meanings of theory: Clarifying theory through typification. *Journal of Management Studies, 58*(2), 1287–1311. https://doi.org/10.1111/joms.12587

Schatzki, T. R. (2002). *The site of the social: A philosophical account of the constitution of social life and change*. Pennsylvania State University Press.

Seastedt, K. P., Schwab, P., O'Brien, Z., Wakida, E., Herrera, K., Grace, P., … Celi, L. A. (2022). Global healthcare fairness: We should be sharing more, not less, data. *PLOS Digital Health, 1*(10), Article e0000102. https://doi.org/10.1371/journal.pdig.0000102

Seibert, K., Domhoff, D., Bruch, D., Schulte-Althoff, M., Fürstenau, D., Biessmann, F., & Wolf-Ostermann, K. (2021). Application scenarios for artificial intelligence in nursing care: Rapid review. *Journal of Medical Internet Research, 23*(11), Article e26522. https://doi.org/10.2196/26522

Seibert, K., Domhoff, D., Fürstenau, D., Biessmann, F., Schulte-Althoff, M., & Wolf-Ostermann, K. (2023). Exploring needs and challenges for AI in nursing care – Results of an explorative sequential mixed methods study. *BMC Digital Health, 1*(1). https://doi.org/10.1186/s44247-023-00015-2. Article 13.

Shabani, M. (2022). Will the European health data space change data sharing rules? *Science, 375*(6587), 1357–1359. https://doi.org/10.1126/science.abn4874

Spradley, J. P. (1979). *The ethnographic interview*. Rinehart and Winston: Holt.

Starren, J., Rasmussen, L., Schneider, D., Nannapaneni, P., & Michelson, K. (2023). A privacy nihilist's perspective on clinical data sharing: Open clinical data sharing is dead, long live the walled garden. *Journal of the Society for Clinical Data Management, 3*(S1). https://doi.org/10.47912/jscdm.234

Susha, I., Rukanova, B., Zuiderwijk, A., Gil-Garcia, J. R., & Gasco Hernandez, M. (2023). Achieving voluntary data sharing in cross sector partnerships: Three partnership models. *Information and Organization, 33*(1), Article 100448. https://doi.org/10.1016/j.infoandorg.2023.100448

Sydow, J., & Windeler, A. (1998). Organizing and evaluating interfirm networks: A structurationist perspective on network processes and effectiveness. *Organization Science, 9*(3), 265–284.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research, 21*(4), 675–687. https://doi.org/10.1287/isre.1100.0323

Vial, G. (2023). Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and Organization, 33*(1), Article 100450. https://doi.org/10.1016/j.infoandorg.2023.100450

Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior results*. Harvard Business School Press.

Weill, P., & Ross, J. W. (2005). A matrixed approach to designing IT governance. *MIT Sloan Management Review, 46*(2), 26–34.

Winter, J. S. (2021). AI in healthcare: Data governance challenges. *Journal of Hospital Management and Health Policy, 5*, 8. https://doi.org/10.21037/jhmhp-2020-ai-05

Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. *The Information Society, 35*(1), 36–51. https://doi.org/10.1080/01972243.2018.1542648

Winter, J. S., & Davidson, E. (2022). Harmonizing regulatory regimes for the governance of patient-generated health data. *Telecommunications Policy, 46*(5), Article 102285. https://doi.org/10.1016/j.telpol.2021.102285

Winter, S., Berente, N., Howison, J., & Butler, B. S. (2014). Beyond the organizational 'container': Conceptualizing 21st century sociotechnical work. *Information and Organization, 24*(4), 250–269. https://doi.org/10.1016/j.infoandorg.2014.10.003

Wixom, B., Sebastian, I., & Gregory, R. (2020). *Data sharing 2.0: New data sharing, new value creation*. MIT Sloan Center for Information Systems Research.

Zeiringer, J. P., & Thalmann, S. (2022). Knowledge sharing and protection in data-centric collaborations: An exploratory study. *Knowledge Management Research & Practice, 20*(3), 436–448. https://doi.org/10.1080/14778238.2021.1978886