5-11-2023

Exploring Design Characteristics of Data Trustees in Healthcare - Taxonomy and Archetypes

Florian Lauf
Fraunhofer Institute for Software and Systems Engineering ISST, florian.lauf@isst.fraunhofer.de

Simon Scheider
Fraunhofer Institute for Software and Systems Engineering ISST, simon.scheider@isst.fraunhofer.de

Jana Friese
Fraunhofer Institute for Software and Systems Engineering ISST, jana.friese@isst.fraunhofer.de

Sarah Kilz
Fraunhofer Center for International Management and Knowledge Economy IMW, sarah.kilz@imw.fraunhofer.de

Marija Radic
Fraunhofer Center for International Management and Knowledge Economy IMW, marija.radic@imw.fraunhofer.de

See next page for additional authors

EXPLORING DESIGN CHARACTERISTICS OF DATA TRUSTEES IN HEALTHCARE — TAXONOMY AND ARCHETYPES

Research Paper

Florian Lauf, Fraunhofer Institute for Software and Systems Engineering ISST, Germany, florian.lauf@isst.fraunhofer.de

Simon Scheider, Fraunhofer Institute for Software and Systems Engineering ISST, and TU Dortmund University, Germany, simon.scheider@isst.fraunhofer.de

Jana Friese, Fraunhofer Institute for Software and Systems Engineering ISST, Germany, jana.friese@isst.fraunhofer.de

Sarah Kilz, Fraunhofer Center for International Management and Knowledge Economy IMW, Germany, sarah.kilz@imw.fraunhofer.de

Marija Radic, Fraunhofer Center for International Management and Knowledge Economy IMW, Germany, marija.radic@imw.fraunhofer.de

Anja Burmann, Fraunhofer Institute for Software and Systems Engineering ISST, and Witten/Herdecke University, Germany, anja.burmann@isst.fraunhofer.de

Abstract

The use of health data provides valuable insights for both research and industry comprising the potential to improve services and facilitate the development of innovative solutions in healthcare. However, due to data protection requirements and technical challenges, access to health data is still severely inhibited. To enhance access to and utilization of health data, science and politics increasingly consider data trustee models as a conceivable solution. Yet, such concepts are still in their infancies and hardly known. At the same time, they exhibit strong differences in their design. Thus, to foster awareness about and the development of data trustee models, this study investigates their design characteristics and integrates them into a holistic taxonomy under the lens of European data law. Additionally, design patterns are explored and archetypes derived. The findings reveal that data trustee models in healthcare follow some overarching design patterns and can be assigned to four dominant archetypes.

Keywords: data trustee, data governance, health data, design characteristics.

1 Introduction

The importance of personal data as a subject of digitization has expanded over the years and nowadays encompasses multiple areas of human life entailing various economic benefits (Leidner and Tona, 2021). In particular, the availability of healthcare data is rising exponentially. Experts estimate that the amount of health related information available worldwide will exceed ten zettabytes by 2025 (Müller, 2021). This includes data about medical histories, diagnoses, treatment suggestions and medical test results as well as data from laboratories, health insurance companies, wearables, and fitness trackers. By sharing and utilizing this information about a person's health status, both research and industry can gain valuable insights and create innovative solutions for improved and more cost-effective patient care,

encompassing personalized medicine and disease control (Kariotis et al., 2020; Marjanovic et al., 2018). In this context, novel digital technologies or platforms can provide a mean for collaboration and enable to address societal challenges. This corresponds to *Goal 3* of the United Nations' *2030 Agenda for Sustainable Development*: "Ensure healthy lives and promote well-being for all at all ages". In Europe, best practices already exist in the public sector. Popular examples are the Finnish health data management and distribution authority *FINDATA* (Specht-Riemenschneider and Kerber, 2022), the cancer registries in Germany (Blankertz and Specht-Riemenschneider, 2021), and organizations such as *MIDATA* (Blasimme et al., 2018; Hafen, 2019) trying to harmonize standards. Besides such early movers, health data are often still unstructured and highly fragmented. Moreover, health data are not yet sufficiently usable to create the expected benefits (Raghupathi and Raghupathi, 2014). In particular, technological barriers and legal obstacles, including data protection and information security, prevent both academic and practice-oriented research from utilizing health data to greater extends (van Panhuis et al., 2014). Though, security is essential to create trust in information systems which is largely related to users' privacy concerns (Anderson, 2000). Information security is defined by ISO/IEC 27000:2009 as the "preservation of confidentiality, integrity and availability of information". Thus, for the purpose of building trust and complying with legal requirements, data trustees must guarantee these security aspects within their technical design. However, specific design characteristics of implementing data security are out of scope. Indisputably, health research in general would benefit in many ways from an adequate data governance model and infrastructure for the exchange and use of sensitive healthcare data (Boyd et al., 2021). For instance, feasibility studies could be conducted in advance of clinical trials to verify sufficient availability of study participants. Other examples for leveraging health related data are effective pattern recognitions of risk factors for diseases or disease progression as well as the evaluation of processes, services and principles of health care (Lesch et al., 2022).

Data governance models for researchers' access to health data is still in its early stages. In this context, the data trustee model is increasingly considered in scientific discussions as a conceivable model. A data trustee represents a fiduciary required to keep the interests of the users at the forefront when making decisions about the processing of their data and ensuring their privacy. In general, such models should enable users to share and utilize their data, while risks to privacy and statutory violation are reliably obviated (Arora, 2019; Bell, 2020; Buchner et al., 2021; Specht-Riemenschneider and Radbruch, 2021). In particular, researchers and companies alike can benefit from cross-organizational data value chains that are commonly established by data trustees in orientation toward the FAIR data principles, i.e., findability, accessibility, interoperability, and reusability (Wilkinson et al., 2016). Resulting, added value can be created for patients, such as approaches to precision medicine, insights into own health data, or support for medical progresses, all while meeting retaining data privacy. While the idea behind data trustees is generally convincing, there is still no common understanding of how a data trustee should be designed to fulfill its intended purpose (Blankertz et al., 2020). However, considerations about data trustees are gradually finding their way into data strategies and current legislation at the European and national levels. In this context, the European Union's Data Governance Act aims at strengthening the exchange and use of data for research and business. It defines and regulates so-called 'data intermediaries' that are supposed to act as independent third parties ensuring trustworthy handling of data. Although various definitions exist, data trustees are commonly described as trusted intermediaries that enable data sharing through a confident and sovereign infrastructure and standardized processes (Lau et al., 2019). However, despite the existence of both theoretical and practical approaches to this emerging concept, design knowledge about data trustees is still scarce. Additionally, a consistent definition and systematic characterization of data trustees has not yet prevailed (Blankertz, 2020). Thus, to remedy the current lack of design knowledge about data trustees, the objective of this paper is to develop a comprehensive taxonomy and, subsequently, inferring archetypes for different trustee models. In our empirical study, we emphasize data trustees in healthcare as we encountered this domain as their predominate application context. Accordingly, we formulate our research questions (RQ):

> **RQ1**: *What are the characterizing elements alongside which data trustee models in healthcare can be structured?*
> **RQ2**: *Which archetypes of data trustee models in healthcare are identifiable?*

To the best of our knowledge, we provide an innovative approach to the implementation and operational realization of data trustees that has not been addressed before. Our research entails broad implications to both theory and practice by identifying general dimensions and characteristics of data trustee models on the one hand and, on the other hand, proposing specific archetypes for their application in healthcare.

The paper is structured as follows. After the *introduction* of Section 1, Section 2 discusses the *theoretical foundation* of data trustees, focusing on the various ways in which data trustees have been designed by former work. Section 3 presents our *research methodology.* The *results* are discussed in Section 4 by means of our developed taxonomy. They are translated into *archetypes* for data trustees in Section 5. The paper ends with a *conclusion* summarizing the main findings, appreciating the limitations, and providing recommendations for future research.

2 Theoretical Foundation

As scientific and political discussions on data trustee models increase, so does the corresponding literature on this novel approach to data governance. We encountered a multitude of publications concerning data sharing between companies and organizations, frequently aiming at making data usable for research purposes. In this section, we clarify our understanding of the term 'data trustee'. Additionally, we provide insights into data trustee models conceptualized by former work.

2.1 Definitional Delimitation

The studies addressing the concept of data trustees commonly highlight the range of parameters that play a role in their operational implementation. However, the existing diversity of design suggestions for data trustees mostly remains on an abstract theoretical level without comprising actually relevant design knowledge. Furthermore, the pluralistic approaches also illustrate that there is no one-size-fits-all solution for a data trustee model and that data trustees require an individual and use case specific structure (Specht-Riemenschneider and Kerber, 2022). The analysis of existing literature further reveals the existence of a plethora of associated terms such as 'data cooperatives', 'data stewardships', or 'data brokers' used synonymously for the concept of 'data trustee'. Likewise, the majority of models tend to differ significantly in their structure, tasks, and characteristics. This results in the fund of literature lacking both a common understanding and a generally accepted definition of the concept.

Many studies deal with data trustee concepts without proposing concrete or operational model approaches. By now, considerations of design characteristics have received little attention in research. A first attempt towards a definition for data trusts is provided by Blankertz and Specht-Riemenschneider (2021). The authors state that three characteristics are generally attributable to a data trust: (1) a function of the data intermediary to manage, distribute, or prepare data for the benefit of others, (2) the objective to fulfill legal requirement in performing activities, and (3) the adherence to application-dependent trust and neutrality requirements in data distribution. Building upon these characteristics, Blankertz and Specht-Riemenschneider (2021) formulate the following definition which is also used as the terminological basis for this research paper [authors translation]:

"A data trustee is a natural or legal person or a business partnership that mediates access to data provided or held by data subjects in accordance with contractually agreed or legally prescribed data governance regulations (also) in the interests of third parties."

2.2 Fundamentals of Data Trustees and Related Work

Following, we present theoretical fundamentals of data trustees as inferable from related work. Building upon evidence gained from three pilot projects, the *Open Data Institute* considers data trustee models as "a legal structure that provides independent stewardship of data" (Hardinges et al., 2019). The authors attribute data trustees the task of taking responsibility (with some obligations) for managing data for an agreed purpose. To this end, the data trustee requires a multidisciplinary team with skills in management,

policy, organizational leadership, data governance, compliance, ethical decision-making, technology, law, user research, economics, product design, and finance (Hardinges et al., 2019).

Reed et al. (2019) provide recommendations for the legal, contractual, and regulatory design of data trustee models. Alike the former work, the author's research is based on experiences from pilot projects. Reed et al. (2019) emphasize that each data trustee needs an individually designed legal form to consider the rights and interests of all potential stakeholders adequately. The authors investigate five different data trustee models from a legal perspective (Reed et al., 2019): the *Traditional Legal Trust Model*, the *Contractual Framework Model*, the *Corporate Model*, the *Public Model*, and the *Community Interest Company Model*. For data sharing between clinics and researchers, Reed et al. (2019) accentuate the Traditional Legal Trust Model as being particularly suitable. In this context, they assume a data trustee model constituted by a small group of data providers and data users, where the parties trust each other and behave appropriately in terms of data sharing and processing (Reed et al., 2019).

As the most concrete approach to conceptualize design knowledge about data trustees, Paprica et al. (2020) focus on the relevant design principles of Canadian data trustees. They involved experts and defined 12 requirements for establishing and operating a data trustee. One of them is that the data trustee must fully comply with all legal requirements. This entails that the design of the data trust model is always strongly affected by the applicable jurisdiction.

Mills (2019) takes a politico-economic perspective on data trusts. The author considers the data trustee model, alongside the laissez-faire approach and the data commons approach, as a model of data ownership concerned with the data flows between the parties involved. Mills (2019) proposes three design options for data trustee models in general: *collector-centric*, *data-centric* and *generator-centric*. In the first model, the trustee acts as an intermediary between users and third parties controlling the flows of data (i.e., collecting data and controlling access). Mills (2019) mentions health data sharing as an example of the collector-centric model. In the data-centric approach, users compile their data into a large data set that is independent of the data service. The trustee negotiates access rights with the data service and receives value in return (e.g., financial, or additional user rights). In the generator-centric model, the data trustee negotiates terms under which users interact with the data collector.

Similar to Mills (2019), Blankertz (2020) discusses the design of data trustees while emphasizing a merely economic perspective. The author argues that data trustees should be designed to enable organizations using data in the best interests of consumers. Blankertz (2020) describes minimum requirements for a data trust to serve as a data steward. These include funding, organizational form, decision-making mechanisms, default setting, and data monetization.

Conclusively, no taxonomy has yet been developed that represents, holistically and comprehensively, the existing dimensions of data trustee models with their characteristics in the healthcare sector. To this end, we develop a taxonomy for data trustee models on which future research can build. Thus, we contribute to the accumulation of design knowledge (e.g., vom Brocke et al. (2020)). Following, we present our research methodology applied to answer our RQs and to remedy the prevailing research gap.

3 Research Methodology

Taxonomies are commonly applied approaches in information systems research to classify, clarify, understand, and systematically examine complex issues (Nickerson et al., 2013). To build our taxonomy, we follow the well-established method of Nickerson et al. (2013). This approach is the de facto standard method to build taxonomies in the field of information systems (Kundisch et al., 2021). As a recently proposed refinement, we incorporated the methodological amendment of Kundisch et al. (2021) adding an evaluation process by means of **focus groups**. The methodological extension of Kundisch et al. (2021) helps to better assess the value of the created taxonomy. Thus, our adapted design process is divided into seven steps that correspond to both the method of Nickerson et al. (2013) and the refinement suggested by Kundisch et al. (2021). Our design process is shown in Figure 1. The shortcuts (**1.**) to (**7.**) refer to the methodological steps amplified in the following.

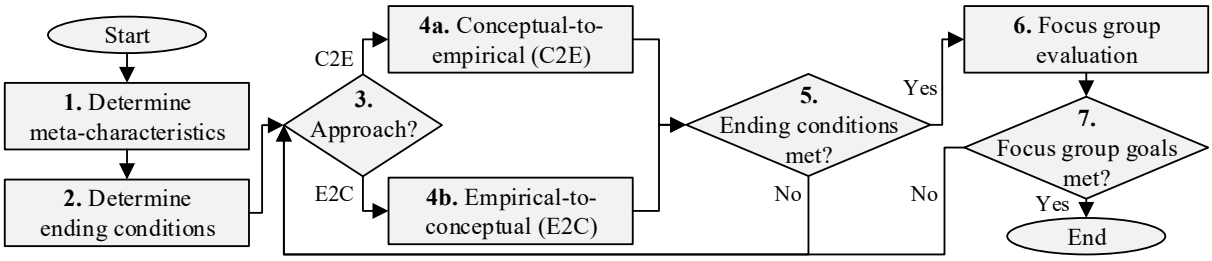

Figure 1. Taxonomy design method adapted from Nickerson et al. (2013) and Kundisch et al. (2021)

Firstly, a meta-characteristic must be specified based on the purpose of the taxonomy so as to each subordinated characteristic and dimension follows from the meta-characteristic (**1.**). Secondly, ending conditions for the iterative part of the process are defined (**2.**). In the subsequent steps (**3.**) to (**5.**), the researchers repeatedly choose between two paths. One path follows a *conceptual-to-empirical* (C2E, **4a.**) approach, in which dimensions and characteristics are inferred from theory. The other path reflects an *empirical-to-conceptual* (E2C, **4b.**) procedure, in which characteristics are inferred from a sample of analysis objects (from practice) and classified in dimensions. After each iteration ((**3.**) to (**5.**)), the ending conditions are checked in step (**5.**) to determine if another iteration is required. If all ending conditions are fulfilled, a further evaluation step (**6.**) follows (Kundisch et al., 2021). We chose focus groups for this adapted evaluation. Focus groups allow to gather more data than individual interviews, since group members do not only put forward their own views on the topics, but rather are required to constantly respond to the input of other group members, triggering discussions and generating new ideas or opinions (Tremblay et al., 2010). According to Szopinski et al. (2019), focus groups are particularly suitable to assess the taxonomy's comprehensiveness, robustness, understandability, and extensibility, as well as the wording of its dimensions and characteristics. The insights from the focus groups regarding any of the criteria can entail changes in form of deletion, alteration, or addition of dimensions and characteristics, which we incorporated into the taxonomy accordingly. In case the results of a focus group discussion (**7.**) do not imply changes of the taxonomy, the latter can be considered finished and the methodological process terminates.

Leveraging the explicated design method, we began by defining the meta-characteristics (**1.**) and, subsequently, determined the ending conditions (**2.**). We directly adopted the ending conditions suggested by Nickerson et al. (2013) (see Table 1). After four design iterations (**4a.**/**4b.**), we considered all ending conditions as fulfilled (**5.**) and initiated the focus group discussion (**6.**), which resulted in minor changes only (**7.**). Thus, after four iterations and one focus group, our taxonomy development was finished. Since we traversed four extensive design iterations and ensured that our experts in the focus groups were knowledgeable in all relevant disciplines (i.e., law, economics, technology), we argue having reached a saturation in results.

In the **1st iteration**, we derived an initial set of dimensions and characteristics from former research in the field (**4b.**), consolidating our definitional delimitation and rated work both addressed in Section 2. Worth mentioning are the works of Specht-Riemenschneider and Kerber (2022), Blankertz (2020), and Paprica et al. (2020). This iteration provided fundamental insights into the possible purposes and activities of data trustees while focusing on the usage of data trustees.

In the following two iterations, we divided our team of authors into (a) technologists and (b) economists / political scientists to ensure objectivity and promoting broader thematical outcomes. This was particularly useful as data trustees are an interdisciplinary field where scientific investigations, especially in terms of design characteristics, require a detailed analysis of technical, economic, and legal concepts. The aim was to provide a comprehensive and holistic analysis of the current state of the art on data trustee models to obtain a broad perspective on the theoretical framework as well as existing best practices, with particular respect to the healthcare sector.

In the **2nd iteration**, our technologists carried out a systematic literature review (SLR) that was based on the databases *Scopus* and *Google Scholar* (**4b.**). We searched for the term 'data trust' to extract

suitable literature, searching for papers published since 2015 and limiting results to the English language. The initial set of results included 631 publications. To develop a comprehensive taxonomy of data trustee models, we did not limit the observation frame to a specific domain or area of application. All papers dealing with data trusts as a tool for data governance were included in the literature review. However, not all publications that we identified met this requirement. Therefore, we filtered the result of the initial search in two phases. Firstly, we checked the publications' title and abstract. Secondly, we assessed the full-text versions regarding the exclusion criteria, removing unsuitable publications. The exclusion criteria were (1) no thematic fit, (2) no access to full-text versions for the authors, (3) duplicates of already included publications within the current iteration. The thematic fit was critical in particular due to the usage of the term 'data trust' in other technical domains (e.g., Internet of Things or wireless sensor networks). The final set consisted of 32 publications. We organized all publications from the final filter phase in a table and analyzed them in terms of new dimensions and characteristics for our taxonomy. We discovered that discussions about the definition of data trustees and their legal classification are prevalent in the existing literature (Stalla-Bourdillon et al., 2021). Interestingly, in this iteration, the data trustee models examined in our final set of literature display different external organizational architectures (e.g., Nabben (2021), Austin and Lie (2021)), internal implementation aspects (e.g., Colliers et al. (2016), Chute et al. (2010), Adhiappan et al. (2020)), and functionalities that provide value for customers (e.g., Wu et al. (2021), Young et al. (2019)).

In the **3rd iteration**, our economists and political scientists carried out an SLR via the *Fraunhofer-Gesellschaft eLib*[1] search portal and the *EBSCO* database while emphasizing their corresponding fields of profession (**4b.**). In the literature search, we applied search terms in German and English, each of which was tested for its relevance in the databases in trial runs. It became evident that the German and English search terms differ or cannot be translated directly, because different words are used in the context of data trustees for the appropriate meaning in the respective languages. The complete search strings were as follows: 1. 'Datentreuhand' OR 'Datentreuhänder' OR 'Datenökonomie' OR 'Datenplattform' OR 'Gesundheitsdaten' AND 'Geschäftsmodell', 2. 'data trustee' OR 'data fiduciary' OR 'data broker' AND 'data intermediaries' OR 'data broker' AND 'data intermediary' OR 'business model' AND 'data intermediaries' OR 'business model' AND 'data intermediary' OR 'business model' AND 'data platform' AND 'health data' OR 'data ownership' AND 'data commons'. The queries in both search portals resulted in a total of 2852 records. We filtered the results in the same steps as in the second iteration. Studies concerned with the development of data governance models, preferably data trustee models, were included. The final set of literature in the 3rd iteration contained 45 articles. This very significant reduction was particularly caused by the term data trust being used in contexts different to data management and governance. Interestingly, despite the different focus, the second SLR did not entail any entirely new dimensions but led to new and merged characteristics of the taxonomy, especially regarding the organizational structure of data trustees (e.g., Falck and Koenen (2020), Micheli et al. (2020), or Delacroix and Montgomery (2020)).

Applying the E2C approach (**4a.**) in the **4th iteration**, we listed existing data trustees that we identified in the first three iterations. We extended this set of analysis objects by a *Google Search* including data trustees from practice not encountered in literature analysis. We searched for German and English data trustees, considering the same search terms used for the first SLR. Although focusing on the healthcare domain, we included data trustees from different application areas in this iteration due to the scarcity of data trustees in practice. In total, we identified 38 data trustees to be considered in this iteration. The assessment of data trustees from practice provided new insights for our taxonomy and a comprehensive fundament for the archetype development. Notably, the new insights changed the taxonomy only slightly (i.e., characteristics and not dimensions). More concretely, the analysis of data trustee models in practice exhibited a decreasing number of small changes to the taxonomy that, ultimately, equalized zero. Since we considered all our ending conditions fulfilled after the fourth iteration (**5.**), we terminated the design iterations and initiated the evaluation phase.

[1] https://fhg-primo.hosted.exlibrisgroup.com/primo-explore/search?vid=FHG_STANDARD02; accessed on Mar 28, 2023

Our taxonomy development method defines that an **evaluation** by means of focus groups is needed to improve and finalize the taxonomy (**6.**). The focus group consisted of ten participants (i.e., authors and additional members) with different spectrums of knowledge ensuring expertise in all relevant fields: technology (4 participants), data economy (3), and data rights (3). We evaluated the taxonomy and discussed our dimensions and characteristics in the plenum while gaining additional insights from the experts. In summary, this evaluation step led to some adjustments in terms of the wording of our taxonomy's dimensions and corresponding characteristics, but did not result in any alterations, additions, or deletions of taxonomy elements. After these minor adjustments, the focus group accepted our taxonomy (**7.**) which we present in Section 4.

Ending Conditions		Design Iterations			
		1	2	3	4
Objective	All papers were examined.			x	x
	No object was merged with another or split into multiple ones.				x
	Each characteristic of every dimension is classified by one object.	x	x	x	x
	No new dimensions or characteristics were added.				x
	Dimensions or characteristics were neither merged nor split.				x
	Each dimension is unique and not duplicated.	x	x	x	x
	Every characteristic is unique within its dimension.	x	x	x	x
	Each cell is unique and not repeated.	x	x	x	x
Subjective	Conciseness – no unnecessary dimensions and characteristics.	x			x
	Robustness – dimensions and characteristics differentiate objects.			x	x
	Comprehensiveness – all objects can be classified.			x	x
	Extendibility – dimensions and characteristics can be added easily.	x	x	x	x
	Explanatory – dimensions and characteristics can describe all objects.			x	x

Table 1. Ending conditions adopted from Nickerson et al. (2013)

4 Results

To answer **RQ1**, we define our meta-characteristic as "*key characteristics of data trustees in the context of healthcare data*". All meta-dimensions, dimensions, and characteristics of our taxonomy are subordinated to this scope. Since we aim at a holistic view on data trustees, we use Leavitt's *diamond model* as meta-dimension (Leavitt, 1965). The model revolves around the four components task, people, structure, and technology; all of them are interconnected and interdependent. Since our taxonomy emphasizes the data trustee concept from both an organizational and a technical design perspective, we resort to the model of Leavitt (1965) to structure our dimensions and characteristics. We merged the components task and people since there are no specific characteristics of human resources relevant for data trustees given our design perspective. However, they are still involved in an operative data trustee model and thus should be recognized somehow in the taxonomy as well. These considerations result in our three meta-dimensions (1) **Task & People**, (2) **Technology**, and (3) **Structure**. In total, they structure 14 dimensions and 44 characteristics. Table 2 shows our final taxonomy, which answers **RQ1**.

We defined the characteristics of the taxonomy mutually exclusive in all dimensions. For the characteristics that are not necessarily mutually exclusive, we added placeholders. Thereof, 'multiple' allows for an appropriate representation of analysis objects. While 'multiple' implies the possibility to select more than one of the listed characteristics, the term 'hybrid' describes a mixed form entailing the appearance of different characteristics. Moreover, 'other' describes additional options not covered in the taxonomy as separate characteristics. An example are cases that are too rare for being included as an individual taxonomy element. The inclusion of such specific cases would lead to a disproportionate expansion of the taxonomy elements. Following, we describe our derived dimensions and characteristics structured alongside the three meta-dimensions inferred from Leavitt (1965).

	Dimension (D_n)	Characteristics (C_{nm})				
Task & People	(D_1) Business Model	C2B	B2B	hybrid		
	(D_2) Main Purpose	data management	self-restraint	conflict resolution	data distribution	
	(D_3) Customer Value	data access	data sharing	data analysis	data quality mngt.	multiple
	(D_4) Data Product	raw data	processed data	analysis results	multiple	
	(D_5) Data Type	personal data	non-personal data	both		
Techn.	(D_6) Core Technology	distributed ledger	other			
	(D_7) Data Integration	static	dynamic	multiple		
	(D_8) Data Storage	centralized	decentralized	hybrid		
Structure	(D_9) Organizational Form	public	private	other		
	(D_{10}) Funding	public financing	private financing	self-financing	hybrid	
	(D_{11}) Deployment Obligation	mandatory	optional			
	(D_{12}) Permitted Scope	no data processing	limited data processing	unlimited data processing		
	(D_{13}) Agency Focus	source-sided	sink-sided	double-sided		
	(D_{14}) Data Altruism	yes	no			

Table 2. Taxonomy of data trustee models in the healthcare sector

4.1 Task & People

According to Leavitt (1965), the meta-dimension 'tasks' describes operations conducted for value-creation in the organization or system under consideration (Saeed and Wang, 2013). In order to perform 'tasks' sufficiently, the organization must implement its chosen data management direction and organizational data capabilities appropriately (Sandrin et al., 2017), referred to as 'people'. Accordingly, the meta-dimension **Task & People** in our taxonomy deals with the work or function of data trustees. Data trustees can be deployed for different tasks and in various domains requiring different *Business Model* (D_1) perspectives (Blankertz et al., 2020). In the healthcare context, data trustees mostly handle patient data, for instance, when patients voluntarily provide their health or health-related data for research (e.g., diagnoses, treatments, or fitness tracker data). Such data can be provided to the data trustee directly by data subjects ($C_{1.1}$) entailing a consumer-to-business (C2B) context. However, most data exchange takes place between business partners ($C_{1.2}$), e.g., clinics or insurance companies representing a business-to-business (B2B) relationship. However, a data trustee does not strictly need to settle for one of these tasks. Thus, hybrid forms are possible ($C_{1.3}$). Related to the business model perspective, data trustees are distinguishable by means of their superordinate purposes (Blankertz and Specht-Riemenschneider, 2021). We defined the dimension *Main Purpose* (D_2) of a data trustee, which we divided into four distinct categories. Firstly, data can be managed to guarantee information security and keep control over data access and use ($C_{2.1}$). Secondly, data trustees can be deployed for self-restraint ($C_{2.2}$) preventing the data user from data protection violations when processing the data. Thirdly, the purpose can be to mediate data exchanges safely without legal or technical risks for any side ($C_{2.3}$). Lastly, data trustees distribute data to facilitate data availability (e.g., for research) while ensuring sufficient data protection ($C_{2.4}$). Naturally, a data trustee can serve multiple purposes, but only one purpose constitutes its main purpose. The purposes of a data trustee are linked to the *Customer Value* provided (D_3) (Bundesdruckerei, 2022a). A first central customer value is data access ($C_{3.1}$). Depending on the perspective, data owners can either prevent illegitimate access to their data, or, in the case of a data user, get access to data by means of the data trustee. Data trustees may also facilitate data sharing for their customers as the main value provided ($C_{3.2}$) or offer a safe platform for such endeavors. Besides data access management, the purpose of data trustees may be to systematically analyze the entrusted

data ($C_{3.3}$) while ensuring data protection during processing. Data trustees can further be used for data quality management ($C_{3.4}$) guaranteeing the entrusted data are valuable for their intended use. Depending on the tasks, a data trustee can provide multiple customer values ($C_{3.5}$). As already touched upon in the dimension *Customer Value*, the kind of actions performed by the data trustee, determines the *Data Product* resulting thereof (D_4) (Blankertz and Specht-Riemenschneider, 2021). Firstly, a data trustee can only manage data access to raw data ($C_{4.1}$). Secondly, data trustees processing entrusted data may provide processed (e.g., standardized) data ($C_{4.2}$) and thirdly, even allow data users to retrieve analysis results ($C_{4.3}$). However, concerning the latter, analysis results are commonly aggregated in a form not allowing an inference to the original content data processed. Fourthly, a data trustee is not limited to one output format ($C_{4.4}$). Furthermore, the data to be managed by a data trustee can cover different *Data Types* which we narrowed down to their personal reference (D_5). Accordingly, we distinguish data trustees handling personal data ($C_{5.1}$), entailing a rigor compliance with legal guidelines (Phillips, 2018), and processing non-personal data ($C_{5.2}$). Naturally, data trustees can deal with both kinds of data types ($C_{5.3}$). However, whenever personal data is the object of investigation, the data trustee must apply sufficient technical and organizational safeguards implemented in its operative system to guarantee their legally correct handling and protection. Concretizations are provided, for instance, in the GDPR.

4.2 Technology

In Leavitt's model, the dimension **Technology** describes how tasks of an organization are performed (Saeed and Wang, 2013). In our taxonomy, we interpret the corresponding meta-dimension as all technical aspects relevant to the implementation of a data trustee. We derived three dimensions with their corresponding characteristics. An important aspect of the technological implementation of a data trustee is the used *Core Technology* underlying the infrastructure (D_6). Many of our sources from both theory and practice did not disclose their particular technology. However, the entirety of models and analysis objects stating a technology leveraged distributed ledger ($C_{6.1}$), specifically blockchain (Nabben, 2021; Zhang, 2021). Since other technologies are conceivable and technology agnostic models exist in the literature, we added a placeholder as a second characteristic ($C_{6.2}$). *Data Integration* (D_7) recognizes the dynamics of the datasets managed by the data trustee. The datasets can be static (i.e., the data do not change, $C_{7.1}$) or dynamic (i.e., the data change after collection, $C_{7.2}$). The latter characteristic is frequently referred to as streaming data. Even though dynamic data exist in healthcare (e.g., data from continuously monitoring devices), such data are usually not per se shared with a data trustee. Finally, the logic of data storing is a relevant aspect entailing an important design characteristic of data trustees (D_8) (Specht-Riemenschneider and Kerber, 2022). In principle, entrusted data can be stored centralized in the data trustee environment ($C_{8.1}$), decentralized in the sphere of the data owners ($C_{8.2}$), or in a hybrid manner ($C_{8.3}$). The latter characteristic addresses the combination of the previous approaches (Specht-Riemenschneider and Kerber, 2022).

4.3 Structure

The organizational **Structure** is used in Leavitt's model to describe the system and location of authority, the communication systems, and the composition of tasks associated with the organizational construct under investigation (Leavitt, 1965; Saeed and Wang, 2013). In our taxonomy, we subordinate design characteristics of data trustees to this dimension describing their organizational aspects. A data trustee can exhibit diverse *Organizational Forms* (D_9). We classify data trustees in this dimension as public ($C_{9.1}$) or private ($C_{9.2}$) organizations (Arlinghaus et al., 2021), while adding the placeholder other ($C_{9.3}$) for hybrid or indistinguishable cases. Similar to D_9, the *Funding* (D_{10}) of data trustees can be provided as public funding ($C_{10.1}$) or by private sponsors ($C_{10.2}$) (Delacroix and Lawrence, 2019). Data trustees may also be self-financed ($C_{10.3}$) collecting membership or usage fees (Blankertz et al., 2020). Additionally, hybrid models ($C_{10.4}$) are possible (e.g., a data trustee funded by public and private sponsors alike). Furthermore, we consider the *Deployment Obligation* of data trustees, referring to the nature of their use being mandatory or optional (D_{11}) (Specht-Riemenschneider and Kerber, 2022). The use of a data trustee is, in most cases, optional for all parties involved ($C_{11.2}$). However, especially in the

healthcare domain, mandatory data trustees exist ($C_{11.1}$) to enable authorized access to sensitive data. Furthermore, as described in the meta-dimension Task & People, a data trustee can assume different tasks. Yet, not all data trustees have the same permissions and thus cannot incur the same tasks (Paprica et al., 2020). This implies that their *Permitted Scope* (D_{12}) in terms of the entrusted data differs. In our taxonomy, we narrowed down the dimension to the characteristics no data processing ($C_{12.1}$), limited data processing (e.g., data analysis) ($C_{12.2}$), and unlimited data processing (e.g., permanently changing or deleting data) ($C_{12.3}$). The scope of permitted activities has direct implications for the *Agency Focus* (D_{13}) of the data trustee (Mills, 2019). When the focus is source-sided ($C_{13.1}$), the data trustee acts as a service provider or broker for data owners. Typical tasks are to protect and manage their data in the best interest. A sink-sided data trustee ($C_{13.2}$) is a service provider for the data users and aims at making data available while data protection is guaranteed. Additionally, there are more balanced settings. An example are two parties intending to share their (sensitive) data with each other. In such contexts, a double-sided data trustee ($C_{13.3}$) is conceivable. Our taxonomy's last dimension is entailed by the Data Governance Act (DGA) and defined by the European Union as data altruism. *Data Altruism* describes the voluntary donation of data for the public good. The aim is to create large datasets for analytics or machine learning. Our homonymous dimension (D_{14}) differentiates data trustees being data altruistic ($C_{14.1}$), thus fulfilling the legal requirements stated in the DGA, or not ($C_{14.2}$).

5 Archetypes of Data Trustees

Our results in Section 4 show that a multitude of different data trustees exists. To answer **RQ2**, we explored patterns in the distinct characteristics of the investigated analysis objects and our literature sample to derive potential archetypes. We searched for (dis-) similarities within cases worth considering by comparing different dimensions of our taxonomy and their corresponding characteristics. Due to the relatively small number of 38 identified data trustees, we opted for a qualitative approach instead of statistical methods such as cluster analysis to extract archetypes from our sample. Specifically, we examined design patterns in our sample of data trustees using the same knowledge base as for taxonomy building. Considering our taxonomy and the data trustee models from practice, we inferred four ideal-typical archetypes, namely **data brokerage trustee**, **data processing trustee**, **data aggregation trustee**, and **data custody trustee**. With respect to the taxonomy's (meta-) dimensions, the archetypes particularly differ in terms of their functions (i.e., tasks) and (external) structure. Dimensions and characteristics related to the technology and internal organization appeared as unsuitable to differentiate data trustees sufficiently. Consequently, the dimensions (D_6) to (D_{11}) are not decisive for our derived archetypes as they depend on concrete use cases. Following, we present our four archetypes, thus providing an answer to **RQ2**. Table 3 points out their corresponding characteristics.

	Dimension (D_n)	Characteristics (C_{nm})				
Task & People	(D_1) Business Model	C2B	B2B		hybrid	
	(D_2) Main Purpose	data management	self-restraint	conflict resolution	data distribution	
	(D_3) Customer Value	data access	data sharing	data analysis	data quality mngt.	multiple
	(D_4) Data Product	raw data	processed data	analysis results	multiple	
	(D_5) Data Type	personal data	non-personal data		both	
Structure	(D_{12}) Permitted Scope	no data processing	limited data processing		unlimited data processing	
	(D_{13}) Agency Focus	source-sided	sink-sided		double-sided	
	(D_{14}) Data Altruism	yes			no	

Legend: Brokerage Trustee ▲, Processing Trustee ■, Aggregation Trustee ●, and Custody Trustee ◆

Table 3. Data trustee archetypes in the healthcare sector

5.1 Data Brokerage Trustee

Data brokerage trustees are neutral intermediaries between data owners and data users. Since this archetype addresses both businesses and consumers, hybrid business models ($C_{1.3}$) are common. Data brokerage trustees serve as mediators protecting the entrusted data and granting access to data users with justified requests. A central obligation of data brokerage trustees is to protect the rights of data owners and support them in controlling both sharing and usage of their data. This implies a source-sided agency focus of the archetype ($C_{13.1}$) since it primarily acts as a representative for the data owner. However, data brokerage trustees are conceivable that are concerned with balancing asymmetries in data exchanges, thus embodying a double-sided agency focus ($C_{13.3}$) (Mills, 2019). Furthermore, by providing supportive data governance, data brokerage trustees can also increase the willingness of data owners to share data. The implementation of effective structures for data governance, data access, and usage control are central tasks for all types of data trustees. In this context, however, data brokerage trustees emphasize an intermediary function for bringing data owners and data users together. Since their focus is on the direct intermediation between data demand and supply side, the main purpose of a data brokerage trustee arises as conflict resolution ($C_{2.3}$) in data exchange. Likewise, they provide added value to customers by means of facilitating data sharing ($C_{3.2}$). Considering a classification in the taxonomy's dimensions Data Product and Data Type, we state data brokerage trustees can offer multiple kinds of data products ($C_{4.4}$) generated from both personal and non-personal data ($C_{5.3}$). We define the permitted scope of data brokerage trustees as limited ($C_{12.2}$) since data owners commonly integrate pre-processed data. In this regard, our investigations in practice have shown that the data brokerage trustee is the most common archetype. Exemplary representatives classifiable in this design pattern are established international data trustee solutions such as *CenTrust* (Bundesdruckerei, 2022a), *MIDATA* (Blasimme et al., 2018; Hafen, 2019), *LunaDNA*, and *Brighthive* (Goodman, 2019).

5.2 Data Processing Trustee

Data processing trustees are characterizable from as business model perspective as actors processing their entrusted data merely in B2B contexts ($C_{1.2}$). The archetype is commonly used as a self-restraint to ensure safe and legally compliant data processing ($C_{2.2}$). Likewise, data processing trustees provide analysis results ($C_{4.3}$) as data products. The archetype is concerned with privacy-preserving data processing, thus avoiding risks to the concealment of the data owners' identities and their data sovereignty. Customer value is provided through data analyses ($C_{3.3}$), i.e., generated data products as referred to above ($C_{4.3}$). Those data products are made accessible to authorized data users, typically in the context of research. An example encountered in the healthcare sector are clinical cancer registries. Specifically, oncologists submit patient health data to the registries that process the data and provide researchers access to the (pseudonymized) results (Bundesdruckerei, 2022b). The provision of analysis results as main purpose inevitably implies that this archetype assumes an unlimited scope of permitted data processing activities ($C_{12.3}$). In principle, data processing trustees can handle both personal and non-personal data ($C_{5.3}$). However, in case of the former, they must obey to the high legal burdens entailed by the GDPR. We define the agency focus of data processing trustees as sink-sided ($C_{13.2}$) due to their emphasis on data utilization (i.e., data product generation) for data consumers. Nevertheless, those data trustees rigorously ensure data privacy of data owners.

5.3 Data Aggregation Trustee

Data aggregation trustees combine data from different owners, typically companies and other institutions ($C_{1.2}$). The archetype creates large, connected datasets linked on an "owner-specific" level. Their focus lies on the aggregation of data from different data owners providing deep insights into and across different fields. This implies that data distribution ($C_{2.4}$) arises as the main purpose of data aggregation trustees. Customer value is provided in multiple ways ($C_{3.5}$), particularly by facilitating data access and ensuring data quality management. Since the collection and linkage of data are frequently combined with the engagement in data analytics services, the archetype is likely to offer multiple data

products which cannot be narrowed down ($C_{3.4}$). This entails a broad scope of data processing activities ($C_{12.3}$), while creating comprehensive datasets from different data owners. As the distribution of enriched and linked data is the main purposes associated with the archetype, data aggregation trustees assume a sink-sided agency focus ($C_{13.2}$). However, they always have to ensure information security when aggregating data. Data trustees encountered in practice which we classified in this archetype usually emphasize the acceleration and improvement of health-related research and patient care by providing data access for authorized researchers. Examples are the *Ontario Health Data Platform*, *FINDATA*, the *Data Integration Centers* used by the *Medical Informatics Initiative Germany* (Medizininformatik-Initiative, 2022), and the *COMET* data trustee (Chan et al., 2021).

5.4 Data Custody Trustee

Data custody trustees assume the primary function of protecting (sensitive) personal data ($C_{5.1}$), while data exchange between companies is paramount ($C_{1.2}$). If at all, data are only made available in anonymized or pseudonymized form. Essentially, the archetype keeps personal identifiable information safe to prevent individuals from being re-identified by any other instance than the data trustee. Thus, the main purpose of this archetype is self-restraint ($C_{2.2}$). For instance, clinics leverage data custody trustees to avoid data protection violations while granting certain entities authorized data accesses. The archetype usually handles highly sensitive personal data for which no legal ground for processing exists ($C_{12.1}$). In line with the given example, this entails that the customer value is merely the enablement of data access ($C_{3.1}$) to raw data ($C_{4.1}$), while prohibiting any forms of processing. The fact that the archetype only handles raw data shows the highly restricted scope of data trustee activities. Since the secure storage and protection of data, including rigor access management, evolves as the primary function of data custody trustees, we define its agency focus as source-sided ($C_{13.1}$). This implies the inconceivability of a data altruistic trustee model ($C_{14.2}$) because data owners' privacy and decisions-makings are in focus. Examples for this archetype encountered in practice are data trustees handling personal data in biobanks. Other application areas are the newborn screening laboratories und screening centers in Germany (Brockow et al., 2022). Different to all other archetypes, the deployment of data custody trustees can be mandatory ($C_{11.1}$) in some healthcare contexts due to the sensitivity of the data under their management.

6 Conclusion and Outlook

The taxonomy detailed in this work provides dimensions and characteristics able to describe and classify data trustees comprehensively. Based on the systematic literature reviews, an analysis of (international) real-world examples, and focus group discussions, we determined 14 dimensions and 44 characteristics to specify data trustee models. To the best of our knowledge, our research represents the first detailed and in-depth work towards investigating the gradually emerging and highly dynamic field of data trustees from an abstract design perspective. The produced insights provide a basic understanding concerning their composition and promote the application of such models to a wide range of interdisciplinary issues. Our research emphasizes the healthcare sector since both theory and practice particularly deal with data trustees in this domain. If required, the applied taxonomy building method allows to dynamically alter both dimensions and characteristics. This is crucial given the rapidly evolving and changing research area, that implies the occurrence and the vanishment of data trustee models with a fast pace. In Europe, as our predominant (legal) lens on the concept, such high dynamics are mainly caused by data protection law (i.e., Data Governance Act). Furthermore, our taxonomy accumulates design knowledge from interdisciplinary disciplines relevant for data trustees (i.e., economics, law, and technology). In this way, the taxonomy enables a broad perspective and deep knowledge of the potential dimensions pertaining to this rapidly evolving concept. Our **scientific contribution** is the interdisciplinary generation of initial design knowledge enabling the implementation of different data trustee scenarios. We derived this knowledge through a theoretically founded methodology from literature and enriched the findings empirically. This design knowledge was aggregated in a taxonomy and a set of archetypes to further understand data trustees and distinguish their characteristics and applications. Scientists can learn from our approach because we merged a

modernized method for taxonomy development with a meaningful combination of research methods. Thereby, we separated our research team to achieve both a higher level of objectivity and a holistic perspective on the interdisciplinary topic. Consequently, it increases scientific rigor of knowledge within this field of high political and legislative attention. Both the taxonomy and the archetypes can be used as an expedient starting point to further develop legislation and organizational endeavors (e.g., European Health Data Space). Most research topics today are at the interface of technology, economics, law, and other disciplines. Our research design ensures that knowledge from those diverse fields is accumulated to create meaningful research efficiently. In terms of **managerial contributions**, our taxonomy enables practitioners to navigate more effectively in the yet mostly unexplored field of designs and compositions of data trustees. Firstly, we offer an overview of the design patterns data trustees are likely to assume, i.e., data brokerage trustee, data processing trustee, data aggregation trustee, and data custody trustee. Secondly, our taxonomy provides practitioners with an interplay of building blocks and prescriptions for effectively designing data trustee models in healthcare. Both the taxonomy and the archetypes help practitioners reflect on their existing approaches and ideate concepts suitable for operation on the one hand. Yet, on the other hand, they are supported to design data trustee models from scratch.

The taxonomy comprises both practical and scientific added value, although it is naturally subject to **limitations**. To some extent, as with all qualitative research, a taxonomy requires an extensive generalization and simplification of most complex issues and their interrelationships (Saldaña, 2015). Moreover, the taxonomy is derived empirically from samples of analysis objects whose representativeness and completeness cannot be certain. Thus, transferability of the results cannot be fully guaranteed and is instead an opportunity for practice-oriented further research. Another limitation is entailed by our taxonomy building process. We mainly relied on published material, which inevitably implies that our results can only build on what is publicly available. Furthermore, subjective opinions and decisions affected our results although we ensured to build our taxonomy and derive our archetypes in teams consisting of authors with different perspectives on the topic, while integrating external experts in focus groups. Moreover, considering the legal and technological dynamics surrounding data trustees, new theoretical models and practical instantiations must be expected to arise soon, while others might disappear with a high frequency. That could result in the need to extend the taxonomy swiftly. To conclude, both our taxonomy and the suggestion of archetypes provide first profound design knowledge about data trustees but require further extension and constant verification.

Future research should focus on a broader validation of our taxonomy and archetypes. By now, the taxonomy can be used as a support tool for developing data trustees in the healthcare sector as it sheds light on the design peculiarities existent in the healthcare domain. We propose future research to apply our findings in other sectors than healthcare and extend them, accordingly. Furthermore, future research should conduct field tests to refine and extend our initial design knowledge about data trustees. Currently, our results compiled in the taxonomy and the suggested archetypes represent initial hypotheses that need validation. Additionally, more data sources (e.g., interviews, case studies, field tests, or literature analyses) should be used, potentially in additional domains, to triangulate a more comprehensive look into the data trustee concept. Future research may also emphasize design science research studies by developing more practice-oriented models and technical instantiations to accumulate more and concretize existing design knowledge about data trustees. We consider future research in the field as crucial since data trustees facilitate a wider access and use of data which enables novel innovations, thus creating benefits for research, industry, and the society. Ultimately, however, the aptitude of this emerging concept to solve the prevailing lack of sharing and utilization data in general and health data in particular is yet to be examined. We state that data trustees can substantially contribute to a fair, trustworthy, and liberal data economy, especially in the highly restricted health domain.

7 Acknowledgement

This research has emerged from the *DaRe* project funded by both the German Federal Ministry of Education and Research (BMBF; funding reference number: 16DTM105A) and the European Union (i.e., NextGenerationEU) as well as the *Digital Life Journey* project funded by Fraunhofer-Gesellschaft.

References

Adhiappan, A., Chernetsov, A., Fenomenov, M., Karabudak, U., Korabanova, A., Kislyakov, S., Beller, L., Nati, M., Radier, B., Sushkov, A., Ustimenko, A., Vedin, A., Yurlov, O., Meriem, T., Messié, V., and Labidurie Omnes, N. 2020. "Federated CSPs Marketplace : A DLT-based Data Trust enabling Business Assurance for CSPs Platforms Federation," Whitepaper, available at https://www.researchgate.net/publication/346630044_Federated_CSPs_Marketplace_A_DLT-based_Data_Trust_enabling_Business_Assurance_for_CSPs_Platforms_Federation, accessed on Nov 16 2022.

Anderson, J. G. 2000. "Security of the distributed electronic patient record: a case-based approach to identifying policy issues," *International journal of medical informatics* (60:2), pp. 111-118 (doi: 10.1016/S1386-5056(00)00110-6).

Arlinghaus, T., Kus, K., Kajüter, P., and Teuteberg, F. 2021. "Datentreuhandstellen gestalten: Status quo und Perspektiven für Geschäftsmodelle," *HMD Praxis der Wirtschaftsinformatik* (58:3), pp. 565-579 (doi: 10.1365/s40702-021-00727-x).

Arora, C. 2019. "Digital health fiduciaries: protecting user privacy when sharing health data," *Ethics and Information Technology* (21:3), pp. 181-196 (doi: 10.1007/s10676-019-09499-x).

Austin, L., and Lie, D. 2021. "Data Trusts and the Governance of Smart Environments: Lessons from the Failure of Sidewalk Labs' Urban Data Trust," *Surveillance & Society* (19:2), pp. 255-261 (doi: 10.24908/ss.v19i2.14409).

Bell, J. L. 2020. "Governing Commercial Access to Health Data for Public Benefit: Charity Law Solutions," *Medical law review* (28:2), pp. 247-269 (doi: 10.1093/medlaw/fwz022).

Blankertz, A. 2020. "Designing Data Trusts. Why We Need to Test Consumer Data Trusts Now," Stiftung Neue Verantwortung (ed.), Stiftung Neue Verantwortung.

Blankertz, A., Braunmühl, P. V., Kuzev, P., Richter, F., Richter, H., and Schallbruch, M. 2020. "Datentreuhandmodelle-Themenpapier,"

Blankertz, A., and Specht-Riemenschneider, L. 2021. "Neue Modelle ermöglichen – Regulierung für Datentreuhänder," 16, Heinrich-Böll-Stiftung (ed.), Heinrich-Böll-Stiftung.

Blasimme, A., Vayena, E., and Hafen, E. 2018. "Democratizing Health Research Through Data Cooperatives," *Philosophy & Technology* (31:3), pp. 473-479 (doi: 10.1007/s13347-018-0320-8).

Boyd, M., Zimeta, M., Tennison, J., and Alassow, M. 2021. "Secondary use of health data in Europe," Open Data Institute.

Brockow, I., Blankenstein, O., Ceglarek, U., Ensenauer, R., Fingerhut, R., Gramer, G., Hörster, F., Janzen, N., Klein, J., Lankes, E., Lindner, M., Mirtschink, P., Murko, S., Rönicke, S., Röschinger, W., Sommerburg, O., Speckmann, C., Winter, T., and Nennstiel, U. 2022. "Nationaler Screeningreport Deutschland 2020," Deutsche Gesellschaft für Neugeborenenscreening e.V., available at https://www.screening-dgns.de/reports.php, accessed on Mar 27 2023.

Buchner, B., Haber, A. C., Hahn, H. K., Prasser, F., Kusch, H., Sax, U., and Schmidt, C. O. 2021. "Das Modell der Datentreuhand in der medizinischen Forschung," *Datenschutz und Datensicherheit - DuD* (45:12), pp. 806-810 (doi: 10.1007/s11623-021-1534-y).

Bundesdruckerei. 2022a. "Der Datentreuhänder: CenTrust Platform der Bundesdruckerei," Bundesdruckerei GmbH, available at https://www.bundesdruckerei-gmbh.de/files/dokumente/pdf/datenblatt-datentreuhaender.pdf, accessed on Nov 9 2022.

Bundesdruckerei. 2022b. "Produkt: Datentreuhänder," available at https://www.bundesdruckerei-gmbh.de/de/loesungen/datentreuhaender, accessed on Nov 4 2022.

Chan, V., Gherardini, P. F., Krummel, M. F., and Fragiadakis, G. K. 2021. "A "data sharing trust" model for rapid, collaborative science," *Cell* (184:3), pp. 566-570 (doi: 10.1016/j.cell.2021.01.006).

Chute, C. G., Beck, S. A., Fisk, T. B., and Mohr, D. N. 2010. "The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data," *Journal of the American Medical Informatics Association : JAMIA* (17:2), pp. 131-135 (doi: 10.1136/jamia.2009.002691).

Colliers, A., Bartholomeeusen, S., Remmen, R., Coenen, S., Michiels, B., Bastiaens, H., van Royen, P., Verhoeven, V., Holmgren, P., Ruyck, B. de, and Philips, H. 2016. "Improving Care And Research

Electronic Data Trust Antwerp (iCAREdata): a research database of linked data on out-of-hours primary care," *BMC research notes* (9), p. 259 (doi: 10.1186/s13104-016-2055-x).

Delacroix, S., and Lawrence, N. D. 2019. "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance," *International Data Privacy Law* (doi: 10.1093/idpl/ipz014).

Delacroix, S., and Montgomery, J. 2020. "From Research Data Ethics Principles to Practice: Data Trusts as a Governance Tool," *SSRN Electronic Journal* (doi: 10.2139/ssrn.3736090).

Falck, O., and Koenen, J. 2020. "Resource "Data": Economic Benefits of Data Provision," *CESifo Forum* (21), pp. 31-41.

Goodman, D. 2019. "The Use of Data Trusts to Improve Impact (Responsibly)," available at https://medium.com/brighthive/the-use-of-data-trusts-to-improve-impact-responsibly-e04aef5f40d9, accessed on Nov 4 2022.

Hafen, E. 2019. "Personal Data Cooperatives – A New Data Personal Data Cooperatives: A New Data Governance Framework for Data Donations and Precision Health," in *The Ethics of Medical Data Donation*, J. Krutzinna and L. Floridi (eds.), Cham (CH), pp. 141-149.

Hardinges, J., Wells, P., Blandford, A., Tennison, J., and Scott, A. 2019. "Data trusts: lessons from three pilots," Open Data Institute (ed.), available at https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit?usp=sharing.

Kariotis, T., Ball, M. P., Greshake Tzovaras, B., Dennis, S., Sahama, T., Johnston, C., Almond, H., and Borda, A. 2020. "Emerging health data platforms: From individual control to collective data governance," *Data & Policy* (2) (doi: 10.1017/dap.2020.14).

Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., and Szopinski, D. 2021. "An Update for Taxonomy Designers," *Business & Information Systems Engineering* (doi: 10.1007/s12599-021-00723-x).

Lau, J., Penner, J., and Wong, B. 2019. "The Basics of Private and Public Data Trusts," *SSRN Electronic Journal* (doi: 10.2139/ssrn.3458192).

Leavitt, H. J. 1965. "Applied organizational change in industry: Structural, technological and humanistic approaches," *Handbook of organizations*.

Leidner, D. E., and Tona, O. 2021. "The CARE Theory of Dignity Amid Personal Data Digitalization," *MIS Quarterly* (45:1), pp. 343-370 (doi: 10.25300/MISQ/2021/15941).

Lesch, W., Richter, G., and Semler, S. C. 2022. "Daten teilen für die Forschung: Daten teilen für die Forschung: Einstellungen und Perspektiven zur Datenspende in Deutschland," in *Datenreiche Medizin und das Problem der Einwilligung: Ethische, rechtliche und sozialwissenschaftliche Perspektiven*, G. Richter, W. Loh, A. Buyx and S. Graf von Kielmansegg (eds.), Berlin, Heidelberg: Springer Nature, pp. 211-226.

Marjanovic, S., Ghiga, I., Yang, M., and Knack, A. 2018. "Understanding value in health data ecosystems: A review of current evidence and ways forward," *Rand health quarterly* (7:2).

Medizininformatik-Initiative. 2022. "Datenintegrationszentren," available at https://www.medizininformatik-initiative.de/de/konsortien/datenintegrationszentren, accessed on Nov 4 2022.

Micheli, M., Ponti, M., Craglia, M., and Berti Suman, A. 2020. "Emerging models of data governance in the age of datafication," *Big Data & Society* (7:2), 205395172094808 (doi: 10.1177/2053951720948087).

Mills, S. 2019. "Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership," London School of Economics & Political Science (ed.), London School of Economics & Political Science, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3437936, accessed on Oct 2 2022.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and PRISMA Group. 2009. "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *Annals of internal medicine* (151:4), pp. 264-269.

Müller, H. C. 2021. "Gesundheitsdaten – Ausgangspunkt für eine lernende Gesundheitsversorgung," Handelsblatt Research Institute (HRI), sponsored by Roche (ed.), Handelsblatt Research Institute (HRI), available at https://www.handelsblatt.com/downloads/27790440/1/gesundheitsdaten.pdf, accessed on Sep 30 2022.

Nabben, K. 2021. "Decentralised Autonomous Organisations (DAOs) as Data Trusts: A General-purpose Data Governance Framework for Decentralised Data Ownership, Storage, and Utilisation," *SSRN Electronic Journal* (doi: 10.2139/ssrn.4009205).

Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A method for taxonomy development and its application in information systems," *European Journal of Information Systems* (22:3), pp. 336-359 (doi: 10.1057/ejis.2012.26).

Paprica, P. A., Sutherland, E., Smith, A., Brudno, M., Cartagena, R. G., Crichlow, M., Courtney, B. K., Loken, C., McGrail, K. M., Ryan, A., Schull, M. J., Thorogood, A., Virtanen, C., and Yang, K. 2020. "Essential requirements for establishing and operating data trusts: practical guidance co-developed by representatives from fifteen canadian organizations and initiatives," *International journal of population data science* (5:1), p. 1353 (doi: 10.23889/ijpds.v5i1.1353).

Phillips, M. 2018. "International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR)," *Human genetics* (137:8), pp. 575-582 (doi: 10.1007/s00439-018-1919-7).

Raghupathi, W., and Raghupathi, V. 2014. "Big data analytics in healthcare: promise and potential," *Health information science and systems* (2), p. 3 (doi: 10.1186/2047-2501-2-3).

Reed, C., Bryan, R., Parker, R., Barwell, E., Martin, C., McMillan, A., Hotchin, J., McIntosh, J., Voznick, M., Cameron, S., Fava, L., Marfé, M., Skene, A., Thomas, C., Snape, R., and Cook, C. 2019. "Data trusts: legal and governance considerations," Open Data Institute, available at http://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf, accessed on Nov 9 2022.

Saeed, B. B., and Wang, W. 2013. "Organisational diagnoses: a survey of the literature and proposition of a new diagnostic model," *International Journal of Information Systems and Change Management* (6:3), p. 222 (doi: 10.1504/IJISCM.2013.058328).

Saldaña, J. 2015. *The coding manual for qualitative researchers*, Los Angeles: Sage.

Sandrin, E., Trentin, A., and Forza, C. 2017. "Organizing for Mass Customization: Literature Review and Research Agenda," *International Journal of Industrial Engineering and Management (IJIEM)* (5:4), pp. 159-167.

Specht-Riemenschneider, L., and Kerber, W. 2022. "Designing Data Trustees - A Purpose-Based Approach," Konrad-Adenauer-Stiftung e. V. (ed.), Berlin: Konrad-Adenauer-Stiftung e. V.

Specht-Riemenschneider, L., and Radbruch, A. 2021. "Datennutzung und -schutz in der Medizin: Forschung braucht Daten," *Deutsches Ärzteblatt* (118:27-28), A1358-A1362.

Stalla-Bourdillon, S., Carmichael, L., and Wintour, A. 2021. "Fostering trustworthy data sharing: Establishing data foundations in practice," *Data & Policy* (3) (doi: 10.1017/dap.2020.24).

Szopinski, D., Schoormann, T., and Kundisch, D. 2019. "BECAUSE YOUR TAXONOMY IS WORTH IT: TOWARDS A FRAMEWORK FOR TAXONOMY EVALUATION," in *Proceedings of the 27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Tremblay, M. C., Hevner, A. R., and Berndt, D. J. 2010. "The Use of Focus Groups in Design Science Research," in *Design research in information systems*: *Theory and practice / Alan Hevner, Samir Chatterjee ; forewords by Paul Gray and Carliss Y. Baldwin*, A. R. Hevner and S. Chatterjee (eds.), New York, London: Springer, pp. 121-143 (doi: 10.1007/978-1-4419-5653-8_10).

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heymann, D., and Burke, D. S. 2014. "A systematic review of barriers to data sharing in public health," *BMC public health* (14), p. 1144 (doi: 10.1186/1471-2458-14-1144).

vom Brocke, J., Winter, R., Hevner, A., and Maedche, A. 2020. "Special Issue Editorial –Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space," *Journal of the Association for Information Systems* (21:3), pp. 520-544 (doi: 10.17705/1jais.00611).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A. C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz,

M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. 2016. "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data* (3) (doi: 10.1038/sdata.2016.18).

Wu, D., Verhulst, S. G., Pentland, A., Avila, T., Finch, K., and Gupta, A. 2021. "How data governance technologies can democratize data sharing for community well-being," *Data & Policy* (3) (doi: 10.1017/dap.2021.13).

Young, M., Rodriguez, L., Keller, E., Sun, F., Sa, B., Whittington, J., and Howe, B. 2019. "Beyond Open vs. Closed," in *Proceedings of the Conference on Fairness, Accountability, and Transparency,* Atlanta GA USA. 29 01 2019 31 01 2019, New York, NY, USA: ACM, pp. 191-200 (doi: 10.1145/3287560.3287577).

Zhang, X. 2021. "A commentary of Data trusts in MIT Technology Review 2021," *Fundamental Research* (1:6), pp. 834-835 (doi: 10.1016/j.fmre.2021.11.016).