



Exploring design elements of personal data markets

Simon Scheider^{1,3} · Florian Lauf¹ · Simon Geller¹ · Frederik Möller^{1,2} · Boris Otto^{1,3}

Received: 29 August 2022 / Accepted: 9 March 2023
© The Author(s) 2023

Abstract

Since the emerging information economy relies heavily on data for advancement and growth, data markets have gained increasing attention. However, while global data economies are evolving and data are increasingly shared among organizations in various data ecosystems, marketplaces for personal data (PDMs) exhibited considerable start-up difficulties, which doomed their majority to either fail quickly or to operate in legal grey zones. Apparently, in recent times, novel PDMs have arisen which seem economically and technically viable. The study investigates this “new generation” from both an economic and a technological perspective. Adhering to a rigorous methodology for taxonomy building and evaluation, 18 dimensions and 59 characteristics are presented alongside which these new PDMs can be designed. Additionally, archetypes are derived. The findings reveal that PDMs tend to follow certain design commonalities holding for data markets generally but comprise specific design elements distinguishing them both from conventional data markets and among each other.

Keywords Archetypes · Personal data markets · Data monetarization · Taxonomy

JEL Classification A1

Responsible Editor: Juho Lindman

✉ Simon Scheider
simon.scheider@isst.fraunhofer.de;
simon.scheider@tu-dortmund.de

Florian Lauf
florian.lauf@isst.fraunhofer.de

Simon Geller
simon.geller@isst.fraunhofer.de

Frederik Möller
frederik.moeller@isst.fraunhofer.de;
frederik.moeller@tu-braunschweig.de

Boris Otto
boris.otto@isst.fraunhofer.de; boris.otto@tu-dortmund.de

¹ Fraunhofer Institute for Software and Systems Engineering, Speicherstraße 6, 44147 Dortmund, Germany

² Chair for Information Systems and Data-Driven Enterprise, TU Braunschweig, Rebenring 58, 38092 Braunschweig, Germany

³ Chair for Industrial Information Management, Dortmund Technical University, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, Germany

Introduction

In recent years, society is witnessing the transition to an information economy building upon data as its raw material, making it the foundation to prosper economically (Parra-Arnau, 2018). Fruhwirth et al. (2020) state that the accruing data-driven business requires market mechanisms and infrastructure for exchange to acquire and monetize valuable data resources. Thus, in the era of the information economy, data, in general, have become a key production factor tradable among different parties. In particular, personal data (PD) represent a central monetary value for the data-driven economy (Yang & Xing, 2019). Spiekermann et al. (2015) distinguish “data assets” between non-personal information and assets consisting of PD. PD have strategic potential as they can lead to superior market intelligence and the improvement of existing operations materializing in enhanced product development (Acquisti & Varian, 2005). Accumulating and using PD can even create competitive advantages for companies (Shapiro & Varian, 1998). Hence, PD are now established as a new kind of asset class, resulting in the emergence of systems dedicated towards their systematic collection, sharing, monetization, and processing. In this context, data markets have arisen as a novel medium to ensure legally effective, economically rational, and technically feasible access to data (Spiekermann & Novotny, 2015; Spiekermann et al., 2015).

Yang and Xing (2019) state the importance of this concept for the safe and effective supply of PD and propose models for personal data markets (PDM), which might solve the existing contradictions between data supply and demand. Spiekermann et al. (2015) address the opportunities and problems of PDMs, contributing to the controversial discourse about the variety of implications arising from trading in PD. Specifically, the authors discuss their economic, technical, legal, social, and ethical challenges. Even though one can observe an increasing relevance of PD in research and practice (Leidner & Tona, 2021; Parra-Arnau, 2018), PDMs are still struggling more than markets trading in non-personal data and must consider this array of issues and their plurality. For example, European PDMs must comply with data protection principles defined in Art. 5 GDPR. This brings a high degree of complexity to the platform architecture (e.g., ensuring appropriation of data processing, data traceability, and security) and the integrated data processing techniques (e.g., privacy-preserving computation). These challenges must, in turn, be combined with economic incentives (e.g., effective data pricing mechanisms) to enable system and business model viability. The struggle becomes evident as many PDMs have disappeared over the past years (e.g., *Datafairplay*, *Datatrade*, *Datawallet*, *MYBS*, *OSA Decentralized*, or *Wibson*). However, despite the problem-sphere outlined above, a new generation of PDMs is currently emerging, which seemingly meets these challenges and is potentially technically and economically viable (Parra-Arnau, 2018). Summarizing, the importance of PD for our information economy is unquestioned (e.g., Leidner and Tona (2021), Spiekermann et al. (2015)), and PDMs are considered a potentially effective facilitator for PD exchange between data providers and consumers (Bruschi et al., 2020).

However, based on our thorough examination of existing literature (see [Theoretical foundations](#) and [Research design](#)), we argue that research lacks conceptually and empirically grounded studies concerning the economic (i.e., business model) and technical (i.e., architecture) design elements of PDMs. We define *design elements* as objects describing the composition of PDMs from an abstract design perspective (i.e., taxonomy characteristics and dimensions). A classification of conjugate design elements is an *archetype* revealing a PDM configuration. Our concrete *research problem* is that existing studies either investigate data markets generally (see [Related taxonomy and literature gap](#)) or develop use case-specific PDM models (e.g., Bataineh et al. (2020), Bruschi et al. (2020), Oh et al. (2019)). Yet, they have neglected the generation of universally valid design knowledge about PDMs, which could aid their development and understanding. Because of the above, the first research question (RQ) reads as follows:

- RQ1: *What are the design elements to structure personal data markets from an economic and technical perspective?*

A taxonomy is a suitable approach to address RQ1 since this artifact provides a set of elementary building blocks and prescriptions for effectively designing an object (Kundisch et al., 2022; Nickerson et al., 2013). We consider the inference of archetypes from our taxonomy to be purposeful, as they are a *basic human mechanism for organizing, summarizing, and generalizing information about the world* (Souza et al., 2007, p. 2). Moreover, archetypes represent foundational conceptual representations, from which manifestations derive, thus accentuating designable (PDM) configurations with certain properties (Möller et al. 2019). Investigating archetypes in poorly researched fields (i.e., PDMs) is particularly useful for fostering a broader understanding of the object under investigation. We propose RQ2 as:

- RQ2: *What are archetypes of personal data markets?*

Our taxonomy consists of dimensions serving to distinguish and explain the design elements of PDMs, while the inferred archetypes give an impression of common PDM configurations. The paper is structured as follows: After the [“Introduction”](#) we continue with the [“Theoretical foundations”](#) to provide a basic understanding of (personal) data markets. We refer to related research to accentuate the added value of our work. In the [“Research design”](#), we explain the methodology for taxonomy development applied to answer RQ1 and RQ2.. The section [“Taxonomy of personal data markets”](#) encompasses both a morphologic box structured alongside five meta-dimensions and the descriptions of its contained design elements. Correspondingly, the section [“Archetypes of personal data markets”](#) amplifies common configurations in which those design elements are likely to occur. In the [“Conclusion”](#) we discuss and summarize our contributions, appreciate the main limitations, and propose future research directions.

Theoretical foundations

Definition of personal data markets

Data markets (DMs) increasingly emerge as electronic marketplaces with data representing the economic good traded on the underlying digital platform (Fruhworth et al., 2020). Stahl et al. (2016) describe DMs generally as electronic platforms facilitating data exchange. They represent a neutral intermediary allowing authorized actors to upload and trade their data (Spiekermann, 2019). Since both industry and research face a rising need to obtain appropriate data to promote innovation and new business potential, the popularity of DMs has grown in recent years (Spiekermann, 2019; Fruhwirth et al., 2020). Likewise, the number of DMs joining the global data market increases constantly. Hence, data

consumers have more possibilities to acquire external data to improve their business, explore new revenue opportunities, and foster innovation and development (Spiekermann, 2019).

Among other authors, Fruhwirth et al. (2020) distinguish personal data markets as a subset of DMs explicitly trading in PD. Since the term PDM is applied to many platforms, networks, and business models, we define the following three characteristics a system must fulfill in our research to be classified as a PDM (e.g., Bataineh et al. (2020), Bruschi et al. (2020), Scheider et al. (2023)): Firstly, a *monetization function* must exist, enabling data providers to yield monetary compensation for disclosing PD, whereby free services are insufficient. Secondly, the system might act as a *direct or indirect intermediary* matching data supply and demand. The actual goods traded on the marketplace can be represented by either data or derivatives resulting from data processing activities performed by or in the marketplace. Thirdly, “people data” must be explicitly considered. Systems trading in anonymized people data are included unless the individuals providing the data have neither access to the system nor any other possibility for direct interaction. The restriction ensures the exclusion of DMs that are entirely uncoupled from the B2C perspective, while hybrid systems or providers trading in anonymized people data are not per se excluded. This characterization based on pertinent literature already suggests that PDMs, albeit commonly considered a subset of DMs (e.g., Fruhwirth et al. (2020)), exhibit significant differences from the latter entailed by the nature of PD as their object of trading. Henceforth, in this study, we refer to DMs as markets trading in non-personal data and thus being mutually exclusive to PDMs trading in personal data.

Personal data being subject to digitalization have expanded over the years and nowadays encompasses multiple areas of human life (Leidner & Tona, 2021). Examples are (1) demographic data (e.g., Hann et al. (2007); Wakefield (2013)); (2) financial data emphasizing transaction and account data (e.g., Dinev and Hart (2006); McKnight et al. (2002)); (3) social media data, in terms of posts, likes, comments, shares, and photos (e.g., Choi et al. (2018); Salehan et al. (2017)); (4) behavioral data capturing daily activities (e.g., Martens et al. (2016)); (5) all kinds of physical actions traced through wearable devices (e.g., Prasopoulou (2017); Warkentin et al. (2017)); and (6) any type of digital traces resulting from user interaction with digital services (e.g., Lehrer et al. (2018)). Many studies attribute monetary value to PD depending on several factors, justifying their treatment as an *economic asset*. In particular, these studies suggest that the pecuniary value rises with the sensitivity of PD (Grossklags & Acquisti, 2007), the entities PD are disclosed to (Cvrcek et al., 2006), and the number of purposes for which PD are provided (Huberman et al., 2005; Spiekermann & Korunovska, 2017). By collecting, processing, and monetizing PD that constantly permeates ever more areas of human existence (Leidner & Tona, 2021), PDMs must be designed in a way supporting

economic viability, while preventing humans from becoming artifacts of technology production (Demetis & Lee, 2018; Spiekermann et al., 2015). This challenge brings to center stage the apparent difference between PDMs and DMs, as the former directly interacts with human beings. In this area of tension, Leidner and Tona (2021) framed the notion of *human dignity* in their CARE theory of PD. It is defined broadly as the recognition that humans possess intrinsic value and, as such, are endowed with certain rights and should be treated with respect. Since humans are represented by their data in digital peripheries and PDMs trade in that data, those markets are subject to high regulation standards (e.g., in Europe, California, and Great Britain) and face ethical responsibilities. To share, process, and monetize such intrinsically valuable PD assets while considering legal and ethical necessities, novel economic mechanisms are crucial and must be embedded in a complex technological architecture. This results in a plurality of interdisciplinary challenges for PDMs (Spiekermann et al., 2015) that do not prevail for DMs. Consequently, PDMs exhibit design peculiarities that are not sufficiently covered by existing DM taxonomies and, as such, have not been examined profoundly yet.

Challenges of personal data markets

Spiekermann et al. (2015) mention a plurality of problems arising for PDMs classifiable into the domains of legal, economic, technical, social, and ethics. The authors argue that such interdisciplinary problems hinder the concept of PDMs from flourishing. For example, in Europe, the legal framework of PD protection, mainly embodied by the GDPR, states a set of restrictive principles like data minimization, legitimate use, purpose binding, and informed consent, leaving little space for market negotiations (Spiekermann et al., 2015). Moreover, even if having designed a legally compliant system, PDMs face economic problems such as that data comprise—in many respects—the traits of free commons (Adams & McCormick, 1987). By their nature, PD are non-rival, cheap to produce, and easy to copy, and can be transmitted without any problems, making PD substantially different from typical commodities. In contrast, taking production data as an example for non-personal data, there is a rivalry among companies participating in a supply chain either trying to exclude others from their usage or engaging in DMs (e.g., Advaneo, DAWEX) for data sharing, monetization, and utilization (Spiekermann, 2019). Furthermore, a legal challenge for PDMs is integrating and tailoring claims to data property. Those claims must be compatible with the notion of privacy as a fundamental right and allow to determine the initial allocation of property rights (Metzger, 2020). Tailoring property rights means restricting alienability and exclusivity of PD (Spiekermann et al., 2015). Beyond economic and legal issues, the interpretation of PD as a tradable economic good requires a philosophical and ethical

discourse about whether human lives, materialized in their data traces, should be property at all or instead be considered inalienable from humans. In this context, some critics from the ethics domain denounce any commercial trading in PD a priori as coercively preclusive with data privacy. This statement predominantly originates from the privacy construct proposed by Solove (2005). In the author's taxonomy, substantial PDM operations (e.g., PD aggregation, tracking, profiling, secondary use, exclusion, and decisional interference) are all recognized as privacy breaches (Spiekermann et al., 2015).

Despite such interdisciplinary challenges accompanying PDMs and a high number having collapsed in practice, a new generation of potentially viable markets seems to be rising in recent years (Parra-Arnau, 2018). Additionally, design-oriented literature increasingly emerges, suggesting many use-case-specific theoretical PDM models (e.g., Bataineh et al. (2020), Bruschi et al. (2020), Oh et al. (2019)). Though, minor advancements have been made in terms of profound examinations of the potentially viable PDMs in practice and pertinent developments in literature.

Related taxonomies and literature gap

Insufficient consideration of PDM design elements becomes evident when analyzing related taxonomies. There are already efforts to describe the design elements of DMs by either taxonomies or typologies. Firstly, Fruhwirth et al. (2020) propose a business model taxonomy based on evidence collected from a sample of 20 DMs. They identified four archetypes of DM business models, one of which is "personal data trading." Our study extends the authors' work by adding an in-depth examination of PDMs since they are too complex and different in design to be considered a DM archetype only (see [Definition of personal data markets](#)). Secondly, Spiekermann (2019) developed a taxonomy consisting of eight differentiating attributes and their corresponding characteristic values, emphasizing the challenges and trends of DMs. Thirdly, Stahl et al. (2016) propose a typology of DMs examining different approaches to data distribution. The authors present a classification framework that outlines a structure for emerging data market research allowing for their distinct definitions. They discern between three ownership types and six corresponding business models (Stahl et al., 2016). Similar to Spiekermann (2019), Stahl et al. (2016) did not differentiate the special case of trading in PD. Lastly, Täuscher and Laudien (2018) provide a taxonomy for marketplace business models by analyzing 100 randomly selected DMs quantitatively. They identified six different marketplace business models that might serve as blueprints for practice. Importantly, as the other studies that disclosed their analysis objects, Täuscher and Laudien (2018) neglected PDMs among their set almost entirely. Conclusively, to the best of our knowledge, existing

taxonomies about data markets do not examine and differentiate PDMs adequately. To address this gap, we constitute the following research design.

Research design

Taxonomy building

Taxonomies are common approaches in information systems (IS) research to classify, clarify, understand, and systematically examine complex issues (Nickerson et al., 2013). We used the taxonomy-building methodology of Nickerson et al. (2013) to identify the design elements of PDMs, as it is frequently used in high-ranking scientific IS outlets. The authors propose generating knowledge conceptually (e.g., from literature) and empirically (e.g., analyzing objects of interest). We supplemented the method with the framework of Szopinski et al. (2019) to complement our design iterations with an evaluation process. Despite this adjustment, our research design still adheres rigorously to the methodology of Nickerson et al. (2013) (see Fig. 1).

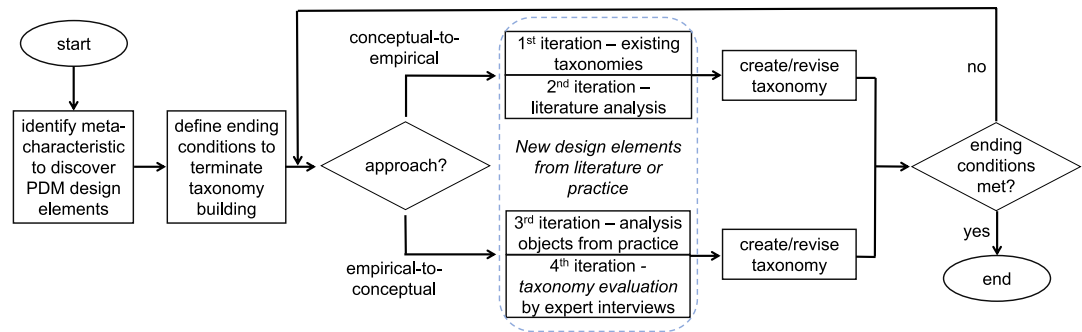
Our research design is dividable into the (iterative) steps shown in Fig. 1. In stages one and two, we chose a meta-characteristic and defined appropriate ending conditions determining the termination criteria of taxonomy development. Each dimension and characteristic must be derived from the meta-characteristic since it is the basis and conceptual border of their discovery. The actual development of dimensions and corresponding characteristics of the taxonomy follows either an *inductive* or a *deductive* approach. In each iteration, the conclusive step checks for the ending conditions. If the criteria for termination are achieved, the taxonomy is finished, and the process terminates. Otherwise, an additional iteration is carried out applying either an inductive or a deductive approach.

Ending conditions, meta-characteristic, and meta-dimensions

We determined our ending conditions following the suggestions of Nickerson et al. (2013) in terms of subjective and objective criteria. Table 1 presents our conditions for termination, together with the iterative taxonomy development. In total, we required four iterations until we considered all conditions of Table 1 fulfilled.

To answer RQ1, we defined our meta-characteristic as "distinguishing key design elements of PDMs from an economic and a technical perspective." This meta-characteristic serves for (1) selecting meta-dimensions, (2) identifying dimensions, and (3) inferring characteristics and their classification to dimensions. To define our meta-dimensions, we relied on the VISOR framework of El Sawy and Pereira (2013). VISOR is adequate for our purpose as it explains how

Fig. 1 The taxonomy development method adapted from Nickerson et al. (2013)



digital platforms (i.e., a PDM) may be designed to respond to a customer need (i.e., demand for PD) profitably and sustainably. To this end, VISOR specifies a set of five partial models (i.e., our meta-dimensions) roughly classifiable in economic (i.e., *value proposition, revenue model, operational model*) and technical aspects (i.e., *service platform, interfaces*). This mixture is inevitable to classify PDMs since both their business models (economic) and architectures (technical) comprise crucial design knowledge that must be unsheathed to remedy the prevailing research gap (see [Introduction](#)). Following, we describe our inductive, and deductive approaches traversed to derive dimensions and characteristics of the taxonomy. After each iteration, we checked the ending conditions of [Table 1](#) until each one was satisfied.

Conceptual-to-empirical design iterations

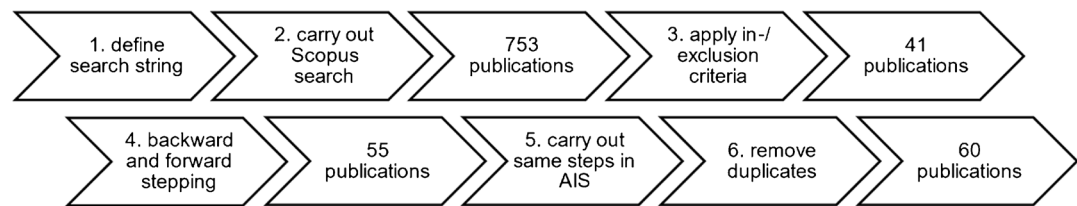
In the 1st iteration, we built an initial taxonomy by merging taxonomies and typologies of former work (see [Related taxonomies and literature gap](#)). In the 2nd

iteration, we carried out an SLR, which is an appropriate strategy to conceptualize dimensions and characteristics (Kundisch et al., 2022) (see [Fig. 2](#)). We follow the method of Kitchenham et al. (2009) (i.e., (1)–(6)), using the RQs of the “[Introduction](#)” as (1) *research questions* guiding the SLR. The (2) *search process* was a manual search for PDM-related conference proceedings and journal papers, including work about DMs. We started by defining our search string and choosing appropriate keywords. For their selection, we drew from pertinent literature (e.g., Fruhwirth et al. (2020)), while ensuring that extracted terms apply to PDMs. We defined the keywords “data marketplace,” “data market,” “trading in personal data,” “personal data market,” “marketplace for personal data,” “data brokers,” “personal data exchange,” “exchanging personal data,” “data monetization,” “monetizing data,” and “data trading.” Those were connected with “OR” operators. Primarily, we searched in Scopus, letting the search engine match our operands in documents’ titles, abstracts, and authors’ keywords. We started with Scopus,

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

Ending Conditions		Design It.			
		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
Subjective	Each cell is unique and not repeated.	X	X	X	X
	Conciseness – no unnecessary dimensions and characteristics			X	X
	Robustness – dimensions and characteristics differentiate objects.			X	X
	Comprehensiveness – all objects can be classified.				X
	Extendibility – dimensions and characteristics can be added easily.				X
	Explanatory – dimensions and characteristics can describe all objects.				X

Fig. 2 Data collection process in the 2nd iteration



as it is an extensive multidisciplinary database covering published material in all fields identified as relevant for PDMs. It also allows for entering precise and long keyword strings (Falagas et al., 2008), particularly important in our research setting. We defined a set of (3) *inclusion and exclusion criteria* to identify and filter relevant papers. Firstly, we excluded the results not available in the English language. Secondly, we included books, albeit trying to avoid them for feasibility reasons. Thirdly, each paper retrieved was reviewed by two authors for whether it addresses the field of PDMs in the broader sense. This means that papers needed to investigate (P)DMs from a technical, economic, ethical, or legal perspective. Due to very broadly formulated keywords in the search string and their combination with “OR” operators, initially retrieved literature contained many papers outside our thematical scope which needed to be excluded. To this end, the third in-/exclusion criterion was examined by screening titles and authors’ keywords before reviewing the entire content of the remaining papers. Since two authors worked together in (3), we argue for reliable objectivity regarding included and excluded papers. Our (4) *data collection* process was determined by the inclusion and exclusion criteria. Their application resulted in 41 papers in our primary Scopus search (see Fig. 2). Building upon those, we conducted backward (referenced articles) and forward (citing articles) stepping (Webster & Watson, 2002), which increased our set by 14 articles. Afterward, we traversed the exact same steps applied for Scopus in AISeL (i.e., (2)–(3)), except that we searched for our keywords in titles only to keep the number of results feasible. Once duplicates were removed, this added five papers to our literature collection. Due to such minor added value, we only carried out “quick searches” in other databases checking whether the top results, firstly, match our in- and exclusion criteria and, secondly, are not already in the collection. Since these quick searches did not add any new papers, we considered the retrieved literature sufficiently representative and terminated our data collection. Finally, our literature set comprised 60 publications, the related taxonomies of the 1st iteration were not part of.

Throughout the steps (2)–(4), we did a (5) *quality assessment*¹ based on the criteria suggested by Kitchenham et al. (2009), i.e., inclusion/exclusion, coverage of relevant articles, quality and validity assessment of literature corpus, and description of the data/studies. For (6) *data analysis*, we extracted phrases (“quotes”) from collected literature with useful content for our taxonomy. Following the approaches of Pratt (2008) and Saldaña (2021), those phrases were coded, included in a system of tables (in Microsoft Excel), and iteratively generalized. As in (4), at least two researchers were engaged in phrase extraction, analysis, and coding to reduce subjectivity biases and find consensus if ambiguity occurred in the process. For a coding example, we refer to our presentation of the “[Taxonomy evaluation](#)” because we used a similar procedure for coding quotes extracted from interviews (4th iteration) and phrases collected from the literature. In the literature analysis, we encountered that much literature about (P)DMs yet exists. The most dominant topics, measured by the number of associated codes, addressed blockchain-based market models. In terms of PDMs, such models frequently apply tokenomics and use smart contracts to structure their organizational and architectural setup. Moreover, legal and ethical discussions in literature related to PDMs commonly encompass aspects of data sovereignty, consent management, and data contracting. We identified many papers about pricing PD as well as (advanced) data security and privacy-preserving computation techniques to ensure data privacy and protection while processing. We state the most important references when delineating our “[Taxonomy of personal data markets](#)”. The SLR also pointed out a lack of studies accumulating, generalizing, and classifying design knowledge about PDMs in a structured manner (see [Introduction](#)).

Empirical-to-conceptual design iteration

In the 3rd iteration, we applied an empirical-to-conceptual approach. We searched for PDMs via the Google search engine using the incognito mode of the browser to

¹ The assessments of these quality criteria are available on request from the authors.

Table 2 Collection of PDMs from practice

PDM	Country	Data Domains	Link
Airbloc Protocol	Korea	holistic	https://www.airbloc.org/
Alga (@Ocean)	British Virgin Islands	undefined	https://datatoken.app/
BIG Token	USA	undefined	https://bigtoken.com/
bitsaboutme	Switzerland	holistic	https://bitsabout.me/de/
Citizen Me	UK	holistic	https://www.citizenme.com/
Clture	USA	video streaming data	https://www.clture.io/
Datacoup	USA	holistic	https://datacoup.com/
Datacy	USA	online activity data	https://datacy.com/consumer
Datareum	UK	holistic	https://www.datareum.net/
Datum	Switzerland	holistic	https://datum.org/
itsmydata	Germany	holistic	https://itsmydata.de/
Kraken	Spain	health and education data	https://krakenh2020.eu/
Madana (Platform)	Germany	holistic	https://www.madana.io/
Permission	Switzerland	holistic	https://permission.io/
PData (opiria.io)	Canada	holistic	https://www.opiria.com/pdata-de/
Polypoly	Germany	holistic	https://polypoly-citizens.eu/de/
Powr of You	UK	online activity data	https://user.powrofyou.com/
Rita Personal Data	Netherlands	holistic	https://ritapersonaldata.com/
Site of Record	no information	business contact data	https://www.siteofrecord.com/
Steamr	Switzerland	undefined	https://steamr.network/
Tartle	USA	holistic	https://tartle.co/
Tiki	USA	undefined	https://mytiki.com/
Vetri	Switzerland	demographic, mobile, and spatial data	https://vetri.global/

circumnavigate carry-over effects from previous searches (Fruhworth et al., 2020). By reusing the keywords from our SLR, we avoided to unconsciously limit our results to PDMs explicitly applying that term in their descriptions. To narrow the results, we leveraged the three characteristics of PDMs defined in our “[Definition of personal data markets](#)” as inclusion criteria. Additionally, we excluded PDMs if it was impossible to obtain meaningful information. We defined the term “meaningful” as access to analyzable information describing the PDM, i.e., technical whitepapers and data retrievable either from analyzing websites or by using (demo) applications (Fruhworth et al., 2020). We also excluded PDMs if the information was meaningful but unavailable in German or English. In contrast to the suggestions of Fruhwirth et al. (2020) for the identification and filtering of DMs, PDMs under construction were not excluded per se. Finally, we obtained 23 PDMs, shown in Table 2. We analyzed these PDMs by classifying them alongside the dimensions and characteristics of our preliminary taxonomy. We searched for information allowing us to assign a PDM to a single characteristic in each dimension. The following three cases could occur: (1) the PDM could be assigned to an existing characteristic in a dimension based on accessible information; (2) the PDM could not be assigned to any characteristic in a dimension due to a lack of accessible

information; (3) the PDM contained information pertaining to a dimension but not associated with a characteristic defined therein. The third case resulted in an alteration or accretion of taxonomy design elements. In the course of the 3rd iteration, the occurrence of this case exhibited a continuous decrease until only (1) and (2) appeared in the PDM analyses. Since the ending conditions were fulfilled at that time, we terminated the taxonomy-building process and initiated the taxonomy evaluation. The set of PDMs was initially created in late 2021 and updated roughly 1 year later. Their analysis was conducted by at least two authors. Despite promising developments in practice (Parra-Arnau, 2018), PDMs still face high dynamics, meaning some analysis objects may vanish while new PDMs are likely to appear.

Taxonomy evaluation

We evaluated our taxonomy in the 4th iteration by means of expert interviews. We chose this method because it allows us to collect data directly on our taxonomy’s design elements and receive concrete feedback. We used the evaluation framework of Szopinski et al. (2019) explicitly tailored to evaluate that artifact class. Table 3 lists our chosen experts. We followed an *expert sampling approach* by selecting industrial and scientific partners from research projects and our personal networks

Table 3 Experts interviewed for taxonomy evaluation

Experts (#I-VII)	Role	Relevant Disciplines (related to (P)DMs)	Time [min]
#I: Science	Research Director	Data Markets/Economics, Technology	47:03
#II: Industry	CEO	Data Markets/Economics, Law	44:38
#III: Science	Research Associate	Data Markets/Economics, Technology	32:12
#IV: Industry	CEO	Data Markets/Ethics	48:55
#V: Industry	Lawyer	Data Markets/Ethics, Law	57:55
#VI: Science	Research Associate	Data Markets/Economics, Technology	57:57
#VII: Industry	Data Scientist	Data Markets/Technology	32:50

(Bhattacharjee, 2012). In particular, we chose our interviewees to ensure expertise about PDMs from both an industry (i.e., practitioners) and a scientific perspective (i.e., researchers). From this initial sample, we invited experts having competencies in the fields relevant to PDMs as postulated by Spiekermann et al. (2015), i.e., technology, ethics, law, and economics. Thus, our finally interviewed experts have both extensive knowledge and sufficiently differentiated points of view on the topic to give comprehensive feedback. Our rough *interview protocol* for semi-structured interviewing was as follows: (1) we presented the research objective and the taxonomy; (2) we explained each dimension with its characteristics and asked the expert for feedback (e.g., in terms of accuracy or usefulness); (3) we asked for design elements missed by the expert in the taxonomy's current version. Essential information about the interviewees can be found in Table 3.

Interviews were conducted by one author, whereas usually, two were responsible for coding and deriving implications for the taxonomy. Because interviews served for evaluation purposes only, and design elements were concretely discussed with experts, we dispensed with detailed coding. Rather, we followed the approach of Pratt (2008), extracting quotes that directly addressed design elements (see Table 4). We generated a code as a highly aggregated imperative of that quote, and sorted imperatives according to the implications (1) no change, (2) delete, (3) add, (4) alter, and (5) rename design element. (1)–(5) represent direct implications for the taxonomy, which were always checked for potential conflicts.

We argue for the minor relevance of subjectivity biases since, firstly, we addressed and discussed specific design elements. Thus, we received concrete feedback leaving hardly any space for false interpretation. Secondly, quotes and codes were generated by at least two authors from transcribed interview material. If their suggestions showed ambiguity, another author was consulted, and quotes, respectively codes, were discussed until reaching consensus. Throughout the interviews, we encountered a remarkable congruence between findings from the first three iterations and expert feedback. In principle, they attributed our taxonomy high comprehensibility, practical relevance, and coverage of important PDM design elements (#I–VII). Concurrently, the interviews led to some minor improvements in terms of the accretion (e.g., #II, #IV–V) and renaming (e.g., #I–II, #V–VI) of characteristics. Notably, the experts only provided feedback on the developed taxonomy, not the phenomenon of PDMs itself. The most significant alteration is partially exemplified in Table 4. Due to rather marginal alterations, and the fact that #VII did not result in any changes, we found a sufficient saturation level after seven interviews. Since we considered the taxonomy complete, we terminated the building process.

Inference of archetypes

Leveraging our taxonomy and analysis results, we identified PDM archetypes in the distribution of PDMs to characteristics across dimensions of our taxonomy. We opted

Table 4 Exemplary quotes and implications for the taxonomy

Exemplary Quotes (author translation)	Code	Implication
#I: “Many such platforms anonymize data [...] and that helps to be compliant with regulation.”	retain <i>anonymization</i> for <i>privacy preservation</i>	no change
#II: “What misses here [...] are all those approaches to make computations over encrypted data [...] I would subsume under the term <i>privacy-preserving computation</i> .”	add <i>privacy-preserving computation</i> to dimension <i>privacy preservation</i>	new characteristic
#III: “We use a kind of multi-party computation tool [...] we have a patent for that [...]”	change <i>privacy-preserving computation</i> to <i>multi-party computation</i>	rename characteristic
#V: “[...] <i>privacy</i> plays a central role here and must be ensured on a technical level. <i>Pseudonymization</i> and <i>anonymization</i> seem like meaningful characteristics.”	retain dimension <i>privacy preservation</i> in its current form	no change

for a qualitative approach instead of statistical methods (e.g., cluster analysis) due to our set of analysis objects (see Table 2) being too small for using quantitative statistics. We used the same knowledge base as for taxonomy building and evaluation, whereas we elaborated on data collected in the 3rd iteration in particular. At least two authors attributed characteristics of dimensions to our “real-world” PDMs based on their corresponding information. A third author was consulted if contradictions occurred. The identification of archetypes followed the methodological guidelines suggested by Yin (2009). The author prescribes recognizing (dis-)similarities within cases and, eventually, separating mutually similar groups between them. This qualitative inference of archetypes was guided by the typology development guidelines of Doty and Glick (1994). Following the authors’ suggestions, we identified influential dimensions that (1) exhibit characteristics clearly observable in the PDMs analyzed and which (2) affect the selection of characteristics in other dimensions. Accordingly, a cluster of characteristics emerged for each PDM. By grouping PDMs with (very) identical characteristics, respectively clusters, we derived three archetypes based on our taxonomy (see [Archetypes of personal data markets](#)). Hence, each archetype exhibits certain configurations ascertained by the characteristics assigned to real-world PDMs in different dimensions.

Taxonomy of personal data markets

Table 5 shows our taxonomy consisting of 18 dimensions (D_n) and 59 characteristics ($C_{n,m}$) spanned over five meta-dimensions adopted from El Sawy and Pereira (2013). We have made adaptations in the selection of dimensions to apply the VISOR framework to our study. Essentially, our meta-dimension “interface” emphasizes data outcome instead of user experience and our “organizational model” focuses on technical platform operation (“processes”) while neglecting the relationships among actors. Considering the nature of our characteristics and the recommendations of Nickerson et al. (2013), we chose mutual exclusiveness. Based on Möller et al., (2020), we argue that the creation of exclusivity does not make our results too complicated due to an additional generalization and linguistic adaptation, still allowing for a clear depiction of each specimen. Furthermore, there are multiple approaches to visualize taxonomies, each entailing different advantages and thus better suited for a specific task (Möller et al., 2022). Following Möller et al. (2022), we consider morphologies as an appropriate visualization as they are concerned with demonstrating the structure and arrangement of taxonomy objects. Following, we describe our taxonomy.

Value proposition

The value proposition emphasizes the particular value created for users by the offering of the PDM (Chesbrough & Rosenbloom, 2002). Additionally, it describes the reduction of focus to a smaller set of potential customers (El Sawy & Pereira, 2013). The dimension *integration* (D_1) addresses the spectrum of data types traded on the PDM. On the one hand, PDMs can offer a comprehensive and rather general variety of data types extended across several domains ($C_{1,1}$) (Dumitru & Gatti, 2016; Hynes et al., 2018). On the other hand, PDMs specialize in one or a few particular data domains ($C_{1,2}$) (Bruschi et al., 2020). *Customer value* (D_2) refers to the PDM’s core offer to users from a high-level perspective. We narrowed down the dimension by defining, firstly, data exchange as the core offer consisting of the switching function of data assets (transaction-centricity) ($C_{2,1}$) and, secondly, data-driven services (data-centricity) (e.g., Spiekermann (2019)). We divided the latter into PDMs selling analysis results as packaged data “products” to data consumers ($C_{2,2}$) without granting them access to PD and PDMs offering “methods” for data processing ($C_{2,3}$) within their (eco-)system. $C_{2,3}$ differs from $C_{2,2}$ in the degrees of freedom the data consumer has in data processing (e.g., combining multiple methods). Methods are typically licensed or bought by data consumers who subsequently process data themselves in a trusted execution environment provided.

Service platform

The meta-dimension service platform maps technology on an organization’s business processes digitally (El Sawy & Pereira, 2013). Likewise, our dimension *primary technology* (D_3) describes the main technological layer the PDM is built upon. Interestingly, the majority of PDMs analyzed leveraged blockchain ($C_{3,1}$) supplemented by smart contracts to integrate processing logic, particularly in, but not limited to, both payment and data exchange transactions. Since a few PDMs use other technologies (e.g., itsmydata, Polypoly) and technology agnostic models exist in the literature, we added a placeholder as second characteristic ($C_{3,2}$). *Infrastructure composition* (D_4) refers to the system architecture as a multi-sided platform. Importantly, it does not address the (de-)central storage of data (as PD have been stored decentral in all PDMs analyzed). Rather, the dimension distinguishes PDMs consisting of a single platform operator ($C_{4,1}$) and ones being distributed among multiple platforms of different service providers ($C_{4,2}$). Furthermore, we define *data input* (D_5) indicating options for transferring external data to a PDM. We identified upload by the user ($C_{5,1}$), user tracking ($C_{5,2}$), and user request ($C_{5,3}$) as differentiable characteristics (Brandão et al., 2019; Shaw & Engels, 2019). A user request is needed in scenarios where the data provider must

Table 5 Taxonomy of personal data markets

Dimension (D_n)		Characteristics (C_{nm})				
Value Proposition	(D_1) Integration	Domain-specific			Domain-unspecific	
	(D_2) Customer Value	Data-as-a-service	Product-as-a-service		Method-as-a-service	
Service Platform	(D_3) Primary Technology	Blockchain and Smart Contracts			Other Technologies	
	(D_4) Infrastructure Composition	Single Operator			Multiple Service Providers	
	(D_5) Data Input	Upload by User	User Tracking	User Request	Multiple	
	(D_6) Data Output	Unprocessed Data	Aggregated Data	Standardized Data	Multiple	
	(D_7) Data Quality Assurance	Reviews by Marketplace		Reviews by Others		None
	(D_8) Processing Activities	All		Limited		Raw Data (None)
Revenue Model	(D_9) Price Discovery	Set by Data Provider	Set by Data Consumer	Market-based Pricing	Hybrid Pricing	
	(D_{10}) Revenue Source	Fee Models	Proceeds of Sale	Others	Multiple	
Operational Model	(D_{11}) Responsibility	Self-Processing			Order-Processing	
	(D_{12}) Consent Management	Specific	Broad	Dynamic	Meta	Others
	(D_{13}) Privacy Preservation	Anonymizing Data	Pseudonymizing Data	Multi Party Computation		Others
	(D_{14}) Token Purpose	Single Token Purpose		Multiple Token Purpose		No Token
	(D_{15}) Currency	Token		Money		Other Cryptocurrency
	(D_{16}) Token Layer	Blockchain (Native)	Protocol (Non-Native)	Application (dApp)		None
Interface	(D_{17}) Output Interface	API	Download	Specialized Software	Multiple	
	(D_{18}) Data Actuality	Static datasets		Streams of data		Multiple

obtain data from a location other than its personal sphere. The characteristics of *tracking* and *upload* describe the situation where PD are already located at the individual and integrated into the PDM manually (static data) or automatized (data streams or static data). Analogously, *data output* (D_6) denominates the possibilities to withdraw purchased values from the PDM, with the output format usually given by dedicated data models. We distinguish unprocessed ($C_{6,1}$), aggregated ($C_{6,2}$), and standardized data ($C_{6,3}$) as basic output types (Perera et al., 2015), whereas $C_{6,2}$ also includes exporting data products. Naturally, combinations are possible ($C_{6,4}$). We define *data quality assurance* (D_7) as mechanisms to guarantee a sufficient quality of data traded, mainly achieved by means of continuously conducted consistency and quality checks. In PDMs, the quality of data is mainly ensured using review systems (Brandão et al., 2019; Travizano et al., 2020). Conceivable characteristics are reviews by marketplace ($C_{7,1}$), reviews by others ($C_{7,2}$), and the absence of any quality assurance ($C_{7,3}$). $C_{7,2}$ encompasses reviews by any actors participating in the PDM other than the (central) PDM operator itself (e.g., data consumers, data providers, or service providers). Finally, *data processing activities* (D_8) denominate value adding operations to data carried out by the PDM. We consider cleansing, standardization (normalization), visualization, cutting, and analysis (Curry, 2016; Koutroumpis et al., 2017). Likewise, we differentiate PDMs offering the entirety of these activities ($C_{8,1}$), a limited number ($C_{8,2}$), or none ($C_{8,3}$).

Revenue model

The revenue model denotes the logic according to which the PDM determines prices and generates income. *Price discovery* (D_9) refers to the PDM's approach to pricing its offerings prior to the transaction (Täuscher & Laudien, 2018), emphasizing the parties being responsible for price setting. Firstly, data providers can set prices ($C_{9,1}$) usually by defining usage or access policies. Secondly, data consumers might state prices ($C_{9,2}$) giving data owners the choice of either accepting or declining an offer. In these two cases, negotiations are possible as data are traded based on individual price-related decisions. Another alternative is a non-negotiation approach, namely market-based pricing ($C_{9,3}$) in which the PDM operator sets the price. In this respect, literature provides much research concerning mathematical models for pricing PD (e.g., auctions). Additionally, hybrid pricing models ($C_{9,4}$) describe data pricing mechanisms affected by decisions of at least two different parties. For example, there are models in which the data provider defines a price. The PDM subsequently alters this price depending on the data quality assessed. The *revenue source* (D_{10}) depicts the income generated by a PDM. PDMs majorly rely on fee models ($C_{10,1}$) like charging a membership fee

or attaching fees to the execution of transactions. However, since PDMs might create customer value through data-based services (see D_2), the generated revenue can also be given by sale proceeds ($C_{10,2}$). Naturally, further revenue sources ($C_{10,3}$) and combinations ($C_{10,4}$) are possible as well.

Operational model

The operational model describes the orchestration of an organization's processes, value chains, and partner relationships to effectively and efficiently deliver its value proposition (El Sawy & Pereira, 2013). Our adapted meta-dimension emphasizes the organizational structure of the PDM. *Responsibility* (D_{11}) refers to the role the PDM assumes in terms of PD processing which (in most jurisdictions) entails the applicability of legal provisions, (e.g., the GDPR in Europe). Basically, PDMs can decide to either self-determine ($C_{11,1}$) PD processing (controller) or only process PD on behalf of data consumers ($C_{11,2}$) (processor). Furthermore, *consent management* (D_{12}) emerges as a central design element of PDMs. Following evidence encountered in literature and practice, we state PDMs can basically leverage specific ($C_{12,1}$), broad ($C_{12,2}$; (Manson, 2020)), dynamic ($C_{12,3}$; (Steinsbekk et al., 2013)), and meta ($C_{12,4}$; (Ploug & Holm, 2016)) consent models. PDMs may even use approaches to blank consent (Angrist, 2009) in very liberal jurisdictions. Since additional forms are also conceivable (e.g., Geller et al. 2022), we add a placeholder ($C_{12,5}$) to cover blank consent models and other special cases. *Privacy preservation* (D_{13}) addresses the technical measures for data privacy applied by the PDM exceeding data encryption. It is frequently considered a vital success factor for PDMs since, in most jurisdictions, privacy-related requirements are mandated by law and support data providers' trust in the system. The dimension addresses the identifiability of individuals providing their data on the PDM. PDMs, in practice, commonly apply techniques of anonymization ($C_{13,1}$) and pseudonymization ($C_{13,2}$). Alternatively, instead of transferring the actual datasets, PDMs can also leverage multi-party computation (MPC) methods ($C_{13,3}$) to share only the results of computations over data while preserving privacy. We added the characteristic others ($C_{13,4}$) indicating for possible further forms of privacy preservation. Examples for $C_{13,4}$ are combinations of the previous three characteristics and the transfer of PD without applying any privacy preservation techniques in addition to encryption. Besides such privacy-related characteristics, PDMs very commonly apply tokenomics in their organizational settings. In this context, the term "token" denominates a piece of data representing a fact or a right within the PDM (Oliveira et al., 2018). We define *token purpose* (D_{14}) emphasizing the organizational and economic perspective of the token used. In a PDM, a token might inherit a single purpose ($C_{14,1}$) like an asset token representing data "ownership," a payment token as an internal means

of payment, a work token distributed as a reward to users incentivizing particular behavior, etcetera. However, PDMs also accumulate several purposes in a single token resulting in the characteristic multiple token purposes ($C_{14.2}$). As we also encountered PDMs not using a tokenomics approach, we define the characteristic of no token ($C_{14.3}$). We added *currency* (C_{15}) to our taxonomy because PDMs require a means of payment. Tokens ($C_{15.1}$) are the payment method most frequently applied, i.e., payment tokens. Though, PDMs also rely on common currencies ($C_{15.2}$) and (other) cryptocurrencies ($C_{15.3}$). Finally, *token layer* (D_{16}) describes the actual integration of tokens in the PDM and thus represents an attribute indicating token implementation and location in a technical sense. We differentiate tokens being native to a blockchain ($C_{16.1}$), issued on top of a protocol ($C_{16.2}$) or located on the PDM's application layer ($C_{16.3}$) (Oliveira et al., 2018). Tokens might run on each layer, albeit the latter appeared most commonly. Naturally, if PDMs do not use any tokenomics concept, no token layer ($C_{16.4}$) is integrated.

Interface

According to El Sawy and Pereira (2013), the success of a company is predicated on the user interface. We adapted the authors' original meta-dimension to include (1) the approaches of PDMs for actually delivering value for users and (2) the actuality of data available on the PDM. *Output Interface* (D_{17}) addresses the possibilities of buyers to access customer value (D_2). Such interfaces are integrated using APIs ($C_{17.1}$), manual file downloads ($C_{17.2}$), specialized software ($C_{17.3}$), or combinations thereof ($C_{17.4}$) (Brandão et al., 2019; Dumitru & Gatti, 2016; Fruhwirth et al., 2020). The dimension does not address whether data and derivatives leave system boundaries or are made accessible within them. It merely describes how purchased items become available for the data consumer. Lastly, we define *data actuality* (D_{18}) paying attention to the kinds of data integrated into the PDM. We differentiate static data ($C_{18.1}$), streaming data ($C_{18.2}$), and multiple ($C_{18.3}$). Streaming data appeared to be most common and are either uploaded continuously or according to fixed schedules. Notably, data input (D_5) emphasizes the alternatives for a data provider to upload data, whereas data actuality focuses on the kind of data capable of being integrated.

Archetypes of personal data markets

We explored archetypes as described in the corresponding section. Since it was not possible to determine, for each PDM analyzed, a characteristic in every dimension, our archetypes rather represent points of reference with anomalies existing. We faced difficulties inferring meaningful archetypes since we encountered multiple similarities

among PDMs, albeit exceptions exist. We discovered (1) a dominance of Blockchain with smart contracts, (2) an extensive usage of tokenomics concepts, and (3) the decentral storage of PD in the individuals' devices. Such overarching design commonalities influenced the development of archetypes in that we needed to identify *influential dimensions* allowing for a clear distinction of PDM configurations (e.g., customer value, infrastructure composition), while neglecting non-influential ones (e.g., primary technology, data input; see [Inference of archetypes](#)). Table 6 shows our PDM archetypes inferred from the taxonomy with the number of classifiable PDMs from Table 2.

As the most common PDM configuration based on empirical examples, Data Traders deliver customer value by providing legally compliant access to PD with reliable quality and in a processable format. Data Traders offer applications for individuals to ingest PD, e.g., via their mobile devices. Imported PD are made available for data consumers in either anonymized or pseudonymized form to preserve privacy. Besides data clearing and standardization, Data Traders apply a limited scope of activities for data enrichment and processing since those are performed by data consumers. Naturally, Data Traders may also handle raw data. Furthermore, prices are determined by either the data consumers or the marketplace operator. The latter is the case if the PDM actively bundles individuals' standardized PD and sells aggregated data packages instead of merely transmitting consumers' data orders to providers and performing transactions. Revenue for Data Traders results from charging fees. PD are processed exclusively on behalf of data providers and consumers since Data Traders provide trading platforms for data exchange while neither storing nor analyzing PD themselves. Legally compliant PD processing and sharing commonly build on specific consent and purchased data are made accessible to consumers via APIs. As illustrative services, Data Traders, firstly, offer access to real-time data (e.g., via HTTPS connections). Secondly, they ensure end-to-end compliance in data exchange (e.g., via hashed user consent information) while, thirdly, facilitating effective and informed consent management. Examples of Data Traders are *Airbloc Protocol*, *BIG Token*, *bitsaboutme*, *Datacoup*, *Datareum*, *Datum*, *Powr of You*, or *VETRI*.

Analysis Service Providers aggregate and process sets of PD, generating data products. Data products are made available for data consumers via downloads or specialized software, preventing them from accessing the actual datasets. Profit for Analysis Service Providers results from the margin between PD monetization paid to individuals of whom data are collected and sale proceeds of data products received from consumers. Prices are determined by the PDMs themselves, which is natural since Analysis Service Providers generate "new" data products requiring pricing. They self-process data, except if data products are exclusively created on behalf of consumers. Analysis Service

Table 6 Archetypes of personal data markets

Dimension	Data Trader	Analysis Service Provider	Data Market Space
(D ₂) Customer Value	(C _{2,1}) Data-as-a-service	(C _{2,2}) Product-as-a-service	(C _{2,1}) Data- or (C _{2,3}) Method-as-a-service
(D ₄) Infrastructure Composition	(C _{4,1}) Single Operator	(C _{4,1}) Single Operator	(C _{4,2}) Multiple Service Providers
(D ₆) Data Output	(C _{6,1}) Unprocessed, (C _{6,2}) Aggregated or (C _{6,3}) Standardized data	(C _{6,2}) Aggregated data	(C _{6,4}) Multiple
(D ₇) Data Qual. Assurance	(C _{7,1}) Reviews by marketplace	(C _{7,1}) Reviews by marketplace	(C _{7,2}) Reviews by others
(D ₈) Data Processing Act.	(C _{8,2}) Limited or (C _{8,3}) None	(C _{8,1}) All	(C _{8,1}) All
(D ₉) Price Discovery	(C _{9,2}) Set by consumer or (C _{9,3}) marketplace	(C _{9,3}) Set by marketplace	No specification
(D ₁₀) Revenue Source	(C _{10,1}) Fee model	(C _{10,2}) Proceeds of sale	(C _{10,4}) Multiple
(D ₁₁) Responsibility	(C _{11,2}) Order-processing	(C _{11,2}) Order-processing	(C _{11,1}) Self-processing
(D ₁₁) Consent Mngt	(C _{11,1}) specific	(C _{11,2}) broad	(C _{11,3}) dynamic or (C _{11,4}) meta
(D ₁₃) Privacy Preservation	(C _{13,1}) Anonymization or (C _{13,2}) Pseudonymization	(C _{13,1}) Anonymization	(C _{13,3}) Multi-Party Computation
(D ₁₇) Output Interface	(C _{17,1}) API	(C _{17,2}) Download or (C _{17,3}) Specialized software	(C _{17,3}) Specialized software
(D ₁₈) Data Actuality	(C _{18,2}) Streams of Data	(C _{18,3}) Multiple	(C _{18,3}) Multiple
Number of PDMs	16/21	5/21	<i>Theoretical concept</i>

Providers are responsible for data quality checks and conduct all relevant processing activities for data product generation themselves. Because such PDMs are seldomly interested in identity attributes, they usually anonymize PD to preserve privacy. Data anonymization enables them to, firstly, circumnavigate legal obstacles even if PD are not processed on exclusive behalf of data consumers and, secondly, apply broad consent models easing PD processing. For instance, *CitizenMe* allows individuals to import and sovereignly monetize their PD by conducting personalized surveys. These surveys are designed by data consumers who receive data visualizations, image galleries, interactive charts, and further data products built upon the anonymized user data processed. The case of *CitizenMe* represents a common business model for Analysis Service Providers.

We define Data Market Spaces (DMS) as decentralized PDMs composed by a network of actors. Actors encompass data suppliers and demanders as well as service providers intermediating between them to share, monetize, and utilize PD. The DMS' network character entails an infrastructure distributed among multiple actors. This distinguishes DMS from Data and Analysis Traders whose infrastructures are provided by a single PDM operator. In principle, decentralized data space concepts are very well-known and increasingly emerge in B2B contexts (e.g., Catena-X, Gaia-X, International Data Spaces, Mobility Data Space, Resilience and Sustainability Data Space, Smart Connected Supplier Network). Evidently, they gradually find their way into B2C environments (e.g., SOLID, Koskinen et al. (2019), Scheider et al., 2023), including PDM-related pilot projects (i.e., DMS). In DMS, service providers offer computational

methods deployable by data consumers on purchased or licensed datasets within the distributed network's boundaries. Thereby, MPC techniques assume an accentuated role to preserve privacy while processing decentrally stored PD. Data outputs for consumers are "self-generated" data products gained by executing computational methods on standardized (and potentially aggregated) data. Consequently, a wide range of processing activities is naturally left with the data consumers applying these methods within the DMS. Since data consumers are the actors responsible for PD processing in DMS networks, they are required to satisfy technical and organizational safeguards entailed by the respective jurisdictions (e.g., GDPR, CCPA). Therefore, service providers offer secure runtime and execution environments within a DMS as specialized software delivering generated data products to consumers. They integrate legally compliant output interfaces in the distributed PDM network. Furthermore, legal safeguards may also prescribe how to obtain consent. For DMS, meta and dynamic consent models are most conceivable (Scheider et al. 2023). Currently, the archetype predominantly exists in theory and publicly funded research projects (i.e., Germany and the EU). The latter are pilot projects in stages of conceptual development and prototyping, thus neglected in our PDM collection (see Table 2).

Conclusion

The taxonomy detailed in this work provides 18 dimensions and 59 characteristics to describe PDMs, accumulating design knowledge generated through examinations in theory and

practice. To the best of our knowledge, it represents the first in-depth work towards investigating this controversially discussed and highly dynamic phenomenon from a use-case independent design perspective. The comprehensiveness of description is achieved when also considering our proposal for archetypes.

The taxonomy's scientific contributions intensify previous work on data markets and their application to the field of PDMs. Consequently, some of the design elements conceptualized in this study draw from prior research while further spinning the red paths of development in terms of PD digitalization (e.g., Leidner and Tona (2021)), emphasizing data allocation and exchange (e.g., Spiekermann et al. (2015)). The taxonomy contributes to existing knowledge, as data market taxonomies do not sufficiently consider the design peculiarities of PDMs caused by their interdisciplinary challenges (see [Challenges of personal data markets](#)). The identified design elements represent PDM practices to meet these challenges. For example, PDMs frequently combine blockchain technology with an appropriate consent model for legal compliance. Effective and informed consent management supports purpose binding and, thus, legitimate use of PD while the chosen technology enables transparency and traceability of PD collection, processing, and sharing. Claims to data property (Metzger, 2020) can be emulated by tokens assigning an exchangeable "ownership like right" within the PDM. Furthermore, the taxonomy reveals design elements for data processing and privacy preservation. Essentially, our results show that, on the one hand, PDMs exhibit design commonalities with DMs. These are, in particular, the dominance of blockchain technology, the classifications of their infrastructure composition (i.e., (de-)centralization of actors) (Fruhvirth et al., 2020) as well as design elements related to data management (i.e., data output, output interface, data actuality) (Fruhvirth et al., 2020; Spiekermann, 2019). Yet, on the other hand, significant design discrepancies appeared between PDMs and DMs as well as among PDMs. The former is entailed by PDMs having to (1) always identify the actor responsible for PD trading and processing (i.e., responsibility), (2) reliably standardize unstructured data and determine data quality scores (i.e., processing activities, data quality assurance), (3) manage individuals' consent (i.e., consent management), and (4) ensure privacy preservation while (5) effectively processing PD. In terms of (5), PDMs are majorly distinguishable among each other by their customer value. Similar to DMs, they may trade in data or data-driven analyses. Alternatively, PDMs may emerge as data spaces for (jointly) sharing, monetizing, and utilizing PD in distributed actor networks. However, due to the aforementioned design discrepancies, the actual configurations of these PDM archetypes differ from the ones known about DMs. Hence, our research shows (how) PDMs are tailored designed for integrating humans with their PD.

In terms of managerial contributions, the taxonomy enables practitioners to navigate more effectively in the yet mostly unexplored field of designs and configurations

of PDMs. Our overview of PDM archetypes helps both researchers and practitioners to anchor and communicate their dimensions and characteristics easily. The taxonomy can be used as a support tool for developing future PDMs, whereat the inferred archetypes assume an accentuated role. They allow to classify and distinguish PDMs, thus fostering practitioners' understanding. In particular, the archetypes facilitate common PDM configurations to practitioners depending on their business purpose (i.e., data trading, analysis services, data space). Furthermore, if required, the used method enables an alteration of the taxonomy, which is vital since PDMs represent a rapidly evolving and changing field where new marketplaces vanish and emerge constantly.

The taxonomy comprises both practical and scientific added value, although it is naturally subject to limitations. Firstly, the PDMs we found might only cover a snapshot of what was available at the time, be outdated quickly, and not be conclusive. Similar limitations hold for our experts, who might not have captured the entire range of possible perspectives on PDMs. Moreover, the analyses of PDM whitepapers, applications, and homepages might have been limited in content. Secondly, there are limitations inherent in our research design. As it is with qualitative research, a taxonomy requires a significant generalization and simplification of most complex issues and their interrelationships (Saldaña, 2021). Furthermore, we have derived our results from potentially limited empirical samples and number of publications. Although we have taken countermeasures, these factors might lead to some interpretative biases being incorporated into the results, e.g., in extracting what each PDM does from public data. Thirdly, extrapolating the trend of the past years, new PDMs must be expected to join the global market quickly, while others might disappear with a high frequency. That implies the future needs to extend the taxonomy swiftly. To conclude, our taxonomy provides first profound design knowledge about PDMs but requires further extension and constant verification, as outlined in our directions for future research.

We advise future research to increasingly investigate PDMs in practice (e.g., by carrying out in-depth case studies). Paying further attention to the challenges of PDMs, future research should also emphasize the development of both conceptual and technological solutions overcoming interdisciplinary obstacles arising for PDMs in multiple domains. We consider research in PDMs as essential since they might be an effective concept to integrate data sovereign humans in the rising information economy, thus supporting innovation and growth through participative value generation.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acquisti, A., & Varian, H. R. (2005). Conditioning prices on purchase history. *Marketing Science*, 24(3), 367–381. <https://doi.org/10.1287/mksc.1040.0103>
- Adams, R. D., & McCormick, K. (1987). Private goods, club goods, and public goods as a continuum. *Review of Social Economy*, 45(2), 192–199. <https://doi.org/10.1080/00346768700000025>
- Angrist, M. (2009). Eyes wide open: The personal genome project, citizen science and veracity in informed consent. *Personalized Medicine*, 6(6), 691–699. <https://doi.org/10.2217/pme.09.48>
- Bataineh, A. S., Mizouni, R., Bentahar, J., & El Barachi, M. (2020). Toward monetizing personal data: A two-sided market analysis. *Future Generation Computer Systems*, 111, 435–459. <https://doi.org/10.1016/j.future.2019.11.009>
- Bhattacharjee, A. (2012). Social science research: Principles, methods, and practices. Textbooks Collection. 3, 2nd edn. University of South Florida. https://digitalcommons.usf.edu/oa_textbooks/3
- Brandão, H., São Mamede, H., & Gonçalves, R. (2019). Trusted data's marketplace. *World Conference on Information Systems and Technologies*, 515–527. https://doi.org/10.1007/978-3-030-16181-1_49
- Bruschi, F., Rana, V., Pagani, A., & Sciuto, D. (2020). Acknowledging value of personal information: A privacy aware data market for health and social research. *CEUR Workshop Proceedings*(2580). http://ceur-ws.org/vol-2580/dlt_2020_paper_6.pdf
- Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, 11(3), 529–555. <https://doi.org/10.1093/icc/11.3.529>
- Choi, B., Wu, Y., Yu, J., & Land, L. (2018). Love at first sight: The interplay between privacy dispositions and privacy calculus in online social connectivity management. *Journal of the Association for Information Systems*, 19(3), 124–151.
- Curry, E. (2016). *The big data value chain: Definitions, concepts, and theoretical approaches*. New horizons for a data-driven economy. Springer. https://doi.org/10.1007/978-3-319-21569-3_3
- Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006). A study on the value of location privacy. *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, 109–118. <https://dl.acm.org/doi/pdf/10.1145/1179601.1179621>
- Demetis, D., & Lee, A. S. (2018). When humans using the IT artifact becomes IT using the human artifact. *Journal of the Association for Information Systems*, 19(10), 929–952.
- Dinev, T., & Hart, P. (2006). An extended privacy calculus model for E-commerce transactions. *Information Systems Research*, 17(1), 61–80. <https://doi.org/10.1287/isre.1060.0080>
- Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review*, 2(19), 203–251. <https://doi.org/10.5465/amr.1994.9410210748>
- Dumitru, R., & Gatti, S. (2016). Towards a reference architecture for trusted data marketplaces: The credit scoring perspective. *2nd International Conference on Open and Big Data (OBD)*, 95–101. <https://doi.org/10.1109/obd.2016.21>
- El Sawy, O. A., & Pereira, F. (2013). Visor: A unified framework for business modeling in the evolving digital space. *Business Modeling in the Dynamic Digital Space*, 21–35. https://doi.org/10.1007/978-3-642-31765-1_3
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>
- Fruhwrith, M., Rachinger, M., & Prlja, E. (2020). Discovering business models of data marketplaces. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 5738–5747. <https://scholarpace.manoa.hawaii.edu/bitstream/10125/64446/1/0567.pdf>
- Geller, S., Müller, S., Scheider, S., Woopen, C., & Meister, S. (2022). Value-based consent model: A design thinking approach for enabling informed consent in medical data research. *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC)*, pp. 81–92. <https://www.scitepress.org/Papers/2022/108280/108280.pdf>
- Grossklags, J., & Acquisti, A. (2007). When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. *Proceedings of Workshop on the Economics of Information Security*. <https://econinfosec.org/archive/weis2007/papers/66.pdf>
- Hann, I.-H., Hui, K.-L., Lee, S.-Y.T., & Png, I. P. (2007). Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems*, 24(2), 13–42.
- Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *IEEE Security & Privacy*, 3(5), 22–25. <https://doi.org/10.1109/MSP.2005.137>
- Hynes, N., Dao, D., Yan, D., Cheng, R., & Song, D. (2018). A demonstration of sterling. *Proceedings of the VLDB Endowment*, 11(12), 2086–2089. <https://doi.org/10.14778/3229863.3236266>
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Koskinen, J., Knaapi-Junnilla, S., & Rantanen, M. M. (2019). What if we had fair, people-centred data economy ecosystems? *IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation* (pp. 329–334). IEEE. <https://ieeexplore.ieee.org/abstract/document/9060350>
- Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). *The (unfulfilled) potential of data marketplaces* (ETLA Working PapersUR - <https://www.econstor.eu/bitstream/10419/201268/1/ETLAWorking-Papers-53.pdf> No. 53). Helsinki: The Research Institute of the Finnish Economy (ETLA).
- Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2022). An update for taxonomy designers. *Business & Information Systems Engineering*, 64(4), 421–439. <https://doi.org/10.1007/s12599-021-00723-x>
- Lehrer, C., Wieneke, A., vom Brocke, J. A. N., Jung, R., & Seidel, S. (2018). How big data analytics enables service innovation: Materiality, affordance, and the individualization of service. *Journal of Management Information Systems*, 35(2), 424–460. <https://doi.org/10.1080/07421222.2018.1451953>
- Leidner, D. E., & Tona, O. (2021). The CARE theory of dignity amid personal data digitalization. *MIS Quarterly*, 45(1), 343–370. <https://doi.org/10.25300/MISQ/2021/15941>
- Manson, N. C. (2020). The case against meta-consent: Not only do Ploug and Holm not answer it, they make it even stronger. *Journal of Medical Ethics*, 46(9), 627–628. <https://doi.org/10.1136/medethics-2019-105955>

- Martens, D., Provost, F., Clark, J., & de Fortuny, E. J. (2016). Mining massive fine-grained behavior data to improve predictive analytics. *MIS Quarterly*, 40(4), 869–888.
- McKnight, D. H., Vivek, C., & Charles, K. (2002). Developing and validating trust measures for E-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
- Metzger, A. (2020). A market model for personal data: State of play under the new directive on digital content and digital services. In S. Lohsse, R. Schulze, & D. Staudenmayer (Eds.), *Data as Counter-Performance – Contract Law 2.0? Münster Colloquia on EU Law and the Digital Economy V, 2020*. <https://ssrn.com/abstract=3666805>
- Möller, F., Bauhaus, H., Hoffmann, C., Niess, C., & Otto, B. (2019). Archetypes of digital business models in logistics start-ups. *Proceedings of the 27th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/17/
- Möller, F., Stachon, M., Hoffmann, C., Bauhaus, H., & Otto, B. (2020). Data-driven business models in logistics: A taxonomy of optimization and visibility services. *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS)*, pp. 5379–5388. <https://scholarspace.manoa.hawaii.edu/items/81b727ac-61d2-4ac6-af80-8fef4ae09e9d>
- Möller, F., Stachon, M., Azkan, C., Schoormann, T., & Otto, B. (2022). Designing business model taxonomies—synthesis and guidance from information systems research. *Electronic Markets*, 32(2), 701–726. <https://doi.org/10.1007/s12525-021-00507-x>
- Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359. <https://doi.org/10.1057/ejis.2012.26>
- Oh, H., Park, S., Lee, G. M., Heo, H., & Choi, J. K. (2019). Personal data trading scheme for data brokers in IoT data marketplaces. *IEEE Access*, 7, 40120–40132. <https://doi.org/10.1109/ACCESS.2019.2904248>
- Oliveira, L., Zavolokina, L., Bauer, I., & Schwabe, G. (2018). To token or not to token: Tools for understanding blockchain tokens. *International Conference of Information Systems, San Francisco, USA*. Advance online publication. <https://doi.org/10.5167/uzh-157908>
- Parra-Arnau, J. (2018). Optimized, direct sale of privacy in personal data marketplaces. *Information Sciences*, 424, 354–384. <https://doi.org/10.1016/j.ins.2017.10.009>
- Perera, C., Ranjan, R., & Wang, L. (2015). End-to-end privacy for open big data markets. *IEEE Cloud Computing*, 2(4), 44–53. <https://doi.org/10.1109/mcc.2015.78>
- Ploug, T., & Holm, S. (2016). Meta consent—a flexible solution to the problem of secondary use of health data. *Bioethics*, 30(9), 721–732. <https://doi.org/10.1111/bioe.12286>
- Prasopoulou, E. (2017). A half-moon on my skin: A memoir on life with an activity tracker. *European Journal of Information Systems*, 26(3), 287–297. <https://doi.org/10.1057/s41303-017-0040-7>
- Pratt, M. G. (2008). Fitting oval pegs into round holes. *Organizational Research Methods*, 11(3), 481–509. <https://doi.org/10.1177/1094428107303349>
- Saldaña, J. (2021). *The coding manual for qualitative researchers* (p. 440). London: SAGE Publications Limited. <http://digital.casalini.it/9781529755992>
- Salehan, M., Kim, D. J., & Changsu, K. (2017). Use of online social networking services from a theoretical perspective of the motivation-participation-performance framework. *Journal of the Association for Information Systems*, 18(2), 141–172. <https://doi.org/10.17705/1jais.00449>
- Scheider, S., Lauf, F., & Geller, S. (2023). Data sovereign humans and the information economy: Towards Design Principles for Human Centric B2C data ecosystems. *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS)*, pp. 3725–3734. <https://scholarspace.manoa.hawaii.edu/items/bae9367e-2959-48d9-8e22-06d714d0bfc>
- Shapiro, C., & Varian, H. (1998). *Information rules: A strategic guide to the network economy*. Harvard Business Press. <https://kcmitt.edu.np/Uploads/information-rulesLarge20210211052224.pdf>
- Shaw, D., & Engels, D. W. (2019). The data market: A proposal to control data about you. *SMU Data Science Review*, 2(3), 13.
- Solove, D. J. (2005). A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477–560. <https://doi.org/10.2307/40041279>
- Souza, E., Lencastre, M., Melo, R. C. F., Ramires, L., & Alves, K. (2007). Analyzing problem frames together with solution patterns. *Proceedings of the 10th Workshop on Requirements Engineering* (pp. 179–189). http://www.inf.puc-rio.br/wer/WERpapers/artigos/artigos_WER07/Twer07-souza.pdf
- Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, 54(4), 208–216. <https://doi.org/10.1007/s10272-019-0826-z>
- Spiekermann, S., & Korunovska, J. (2017). Towards a value theory for personal data. *Journal of Information Technology*, 32(1), 62–84. <https://doi.org/10.1057/jit.2016.4>
- Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, 25(2), 161–167. <https://doi.org/10.1007/s12525-015-0191-0>
- Spiekermann, S., & Novotny, A. (2015). A vision for global privacy bridges: Technical and legal measures for international data markets. *Computer Law & Security Review*, 31(2), 181–200. <https://doi.org/10.1016/j.clsr.2015.01.009>
- Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). A classification framework for data marketplaces. *Vietnam Journal of Computer Science*, 3(3), 137–143. <https://doi.org/10.1007/s40595-016-0064-2>
- Steinsbekk, K. S., KåreMyskja, B., & Solberg, B. (2013). Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem? *European Journal of Human Genetics*, 21(9), 897–902. <https://doi.org/10.1038/ejhg.2012.282>
- Szopinski, D., Schoormann, T., & Kundisch, D. (2019). Because your taxonomy is worth IT: Towards a framework for taxonomy evaluation. *Proceedings of the Twenty-Seventh European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/104/
- Täuscher, K., & Laudien, S. M. (2018). Understanding platform business models: A mixed methods study of marketplaces. *European Management Journal*, 36(3), 319–329. <https://doi.org/10.1016/j.emj.2017.06.005>
- Travizano, M., Sarraute, C., Dolata, M., French, A. M., & Treiblmaier, H. (2020). Wibson: A case study of a decentralized, privacy-preserving data marketplace. *Blockchain and Distributed Ledger Technology Use Cases*, 149–170. https://doi.org/10.1007/978-3-030-44337-5_8
- Wakefield, R. (2013). The influence of user affect in online information disclosure. *The Journal of Strategic Information Systems*, 22(2), 157–174. <https://doi.org/10.1016/j.jsis.2013.01.003>
- Warkentin, M., Goel, S., & Menard, P. (2017). Shared benefits and information privacy: What Determines smart meter technology adoption? *Journal of the Association for Information Systems*, 18(11), 758–786. <https://doi.org/10.17705/1jais.00474>
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), 13–23.
- Yang, J., & Xing, C. (2019). Personal data market optimization pricing model based on privacy level. *Information*, 10(4), 123–138. <https://doi.org/10.3390/info10040123>
- Yin, R. K. (2009). *Case study research: Design and methods* (vol. 5). Los Angeles: SAGE.

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