Data Sovereign Humans and the Information Economy: Towards Design Principles for Human Centric B2C Data Ecosystems

Simon Scheider
Fraunhofer ISST
TU Dortmund University
simon.scheider@isst.fraunhofer.de

Florian Lauf
Fraunhofer ISST
florian.lauf@isst.fraunhofer.de

Simon Geller
Fraunhofer ISST
simon.geller@isst.fraunhofer.de

Abstract

The ever-growing amounts of data offer companies many opportunities for data-driven-value generation which, in turn, can be multiplied by leveraging data across company boundaries in evolving data ecosystems. However, while such systems increasingly emerge in B2B environments enabling systematic sharing and utilization of "industrial data", comparable concepts in B2C ambits have not yet prevailed. Despite the rising importance of personal data in the information economy, B2C data ecosystems represent a widely unexplored research area. To remedy this gap, the study generates design principles for human centric B2C data ecosystems to aid in their development. For this purpose, a qualitative interview study with experts of interdisciplinary domains and a structured literature review are conducted both embedded into a methodology for generating design principles. On this basis, derived design principles help to understand peculiarities of data ecosystems in B2C ambits and provide solutions to overcome their obstacles identified in the empirical investigation.

Keywords: Data Ecosystems, Data Sovereignty, Design Principles, Personal Data.

1. Nature and value of personal data

Personal data (PD) being subject to digitalization has expanded over the years and nowadays encompasses multiple areas of human life while entailing various economic benefits (Leidner & Tona, 2021). In fact, the production, the collection, and the processing of huge quantities of data about humans and their activities offer new opportunities to the information economy (Schwartz, 2004). From the perspective of organizations, leveraging PD can generate enhanced customer "knowledge", either about individuals or homogenous groups of people. These data-processing organizations optimize their operations, innovate new and tune existing products, and improve their overall business processes (Spiekermann et al., 2015). Thus, the availability of richer sets of PD increasingly evolves as the key enabler for innovative products and services produced in constantly shortened development cycles.

Likewise, data in general and PD in particular are increasingly considered as economic assets (Birch et al., 2021). However, there is a discrepancy in value conception, since data comprise relatively low value for humans creating and holding them, but are valuable, especially in the aggregate, for data-processing organizations (Oehler, 2016). This discrepancy results in humans on the supply side of the data economy willingly disclosing their PD for free, while enormous commercial value is created on the demand side without sharing any profit with individuals as the producers of personal data (Lauf et al., 2022). The dissemination and industrialisation of such digital business models have caused the emergence of both contemporary and restrictive frameworks for data protection around the globe to regulate parts of the rapidly evolving data economy (Oehler, 2016). However, they frequently lack an innovation perspective, as it is assumed humans have an interest in data privacy only (Oehler, 2016).

Hummel, Braun, and Dabrock (2021) point out that humans have an interest in sharing their PD given that they obtain a fair share on generated profits. The authors distinguish a protective and a participatory claim to PD which create humans' data sovereignty. Lauf et al. (2022) point towards challenges arising when unifying the divergent concepts of humans' data sovereignty and data economics. Together with the aforementioned false perceptions prevailing in our data economy, pinpointed by, for instance, Hummel, Braun, and Dabrock (2021) and Oehler (2016), such clashes of interest impede innovation and growth. As a result, while in business-to-business (B2B) environments industrial data are increasingly shared and utilized collaboratively in evolving data

URI: https://hdl.handle.net/10125/103087
978-0-9981331-6-4

ecosystems (e.g., Catena-X, Gaia-X, International Data Spaces), there are hardly any developments towards such systems in business-to-consumer (B2C) ambits. This is unfortunate since former work has already pointed out that, albeit being hardly explored (Koskinen et al., 2019), user-centric data ecosystems may be an enabler for PD sharing and utilization while ensuring humans' data sovereignty (Rantanen & Koskinen, 2020; Sambra et al., 2016). Currently, the most known development in practice is SOLID that provided vital information in the course of our study.

Conclusively, research is urgent to further investigate and design the promising concept of human centric B2C data ecosystems. Consequently, our **research question** is: *What are conceptually grounded and empirically validated design principles for human centric B2C data ecosystems?*

Our research contributes to an extension of scientific domain knowledge. Specifically, it provides a solid foundation for academic discourses and future research related to the topics of data sovereignty and the design of data ecosystems in B2C contexts.

The paper continues with the theoretical foundations, firstly, explaining the two-sided concept of data sovereignty and, secondly, introducing data ecosystems. In Section 3, we outline our research design, particularly the methodological process of Möller et al. (2020) to conceptualize design principles (DPs). In Section 4, we present our empirically derived DPs that are subsequently amplified in Section 5. We close with the scientific and managerial contributions of our work, an appreciation of limitations, and directions for future research.

2. Theoretical Foundations

2.1 Data sovereignty for human centricity

The term sovereignty encompasses claims to power and control that are linked to reciprocal concessions and relationships of recognition (Maritain, 1950; Schmitt, 2005). In literature, *data sovereignty* is commonly understood as a subtype of sovereignty addressing the empowerment of actors to exercise control functions over the use of their data (Adonis, 2019; Couture & Toupin, 2019). This comprises controllability of the entities having access to data, determination of allowed purposes under which data may be processed and clarity of how access and processing affect the actors' exercise of freedom (Hummel, Braun, Augsberg et al., 2021).

Hummel, Braun, Augsberg et al. (2021) state that data sovereignty is originally linked to a defence perspective, which concerns the protection of humans' liberties. This perspective addresses *control claims* encompassing the protection of privacy (Véliz, 2020), i.e., the ability to shield data from access and processing. Furthermore, data sovereignty requires that data can be attributed to specific entities who inevitably assume the right to execute such control claims (Hummel, Braun, & Dabrock, 2021). The determination of control, exclusion, and exploitation rights humans can exercise on their PD evolves as a central part of the debate on whether there is an ownership right to data (Lohsse et al., 2020).

From this legal perspective, it is discussed how an original *sui generis* intellectual property right to humans' PD should be designed (Fezer, 2017b). Such a legal entitlement must ensure that humans are provided with an actual right of defence and property to their PD (Fezer, 2017a; Lohsse et al., 2020). These considerations of **protective claims** related to data sovereignty are required to enable humans making enforceable decisions about their data (Hummel, Braun, & Dabrock, 2021). Thus, in line with pertinent literature, we attribute to the control aspect the possibility and empowerment of humans to determine the entities accessing their data (Couture & Toupin, 2019), the purpose for which data are processed (Werthner & van Harmelen, 2017), and the ability to appraise (or observe retrospectively) the consequences arising for humans' privacy (Hummel, Braun, Augsberg et al., 2021). Importantly, data privacy does not simply exist in nowadays information economy, but must rather be fought for, shaped and defended constantly. This requires control over inferences and repercussions from the use, the analysis, and the prediction of personal data (Hummel, Braun, Augsberg et al., 2021; Lauf et al., 2022).

However, data sovereignty is not limited to the adumbrated protective claim. As social and networked beings, humans have an interest in the creation of information flows and their utilization (Hummel, Braun, & Dabrock, 2021; Lauf et al., 2022). Thus, data sovereignty also encompasses **participatory claims** to data (Hummel, Braun, Augsberg et al., 2021). Those claims enable humans to strike a self-determined balance between shielding data (protective claims) and making data available in a controllable manner. If data can be consciously leveraged for specific purposes by means of enforceable participatory claims, humans inevitably participate in data-driven coordination, knowledge and innovation processes of organizations (Lauf et al., 2022; Pohle & Thiel, 2020). This corresponds to the model of a self-determined human in the digital world as postulated by Meister and Otto (2019) providing the theoretical basis for human centric data ecosystems. We consider data sovereignty as the foundational concept for B2C data ecosystems, making those systems human centric.

2.2 Fundamentals of data ecosystems

In their theory of digital ecosystems, Jacobides et al. (2018) describe a digital ecosystem as an interacting organization enabled by modularity and managed without any hierarchical order. The authors emphasize a business aspect stating that the modular endpoints of the ecosystem are bound together by the impossibility of allocating their collective investment elsewhere (Jacobides et al., 2018). Data ecosystems represent a subset of digital ecosystems with the central purpose of sharing and jointly utilizing data (Oliveira et al., 2019).

Digital or data ecosystems are classifiable as open, dynamic and complex networks of actors (Li et al., 2012). Openness entails a "flow of energy" is required to maintain the system state between both the system and its environment as well as different system entities. Digital ecosystems exhibit diverse temporal and spatial scales of dynamic developments and their complexity is determined by the number of interactions between actors (Currie, 2011; Li et al., 2012). Furthermore, three essential characteristics can be attributed to digital ecosystems (Jansen et al., 2013). The *network character* describes them as loosely coupled networks of actors. The *platform character* entails the existence of services, tools or technologies actors can use in the ecosystem for creating value. The characteristic of *co-evolution* addresses actors using the system to create innovations by pooling capabilities and resources developed through their mutual collaboration and connection. The different relationships of actors to resources are the reason for the emergence of roles, which are comparable to functions performed by actors in the ecosystem (Hanssen & Dyba, 2012; Oliveira et al., 2019). Typically, a central function emerges that is predominantly responsible for system viability (Hanssen & Dyba, 2012).

As a subset of digital ecosystems, data ecosystems represent complex socio-technical networks that consist of, firstly, autonomous actors collaboratively utilizing data and, secondly, an environmental setting for creating, managing, and sustaining data sharing initiatives (Oliveira et al., 2019), e.g., smart cities (Abu-Matar, 2016), open data (Lee, 2014), or scientific data communities (Lindman et al., 2015). In order to unlock the potential benefits of data across companies, industries, and even entire countries, data ecosystems are nowadays considered an auspicious medium in our data economy (Oliveira et al., 2019).

The emergence of data ecosystems in B2B ambits is driven by multiple factors, including new digital technologies and political initiatives worldwide, e.g., Gaia-X, open data movement, and open government data programs, which call for the free (re-) use and distribution of data by anyone (Oliveira et al., 2019). Improvements and trends in the technologies are also driving private and public organizations to publish data and to integrate their services with external data.

However, while data ecosystems are arguably gaining in importance, both research in and practical developments of B2C (or C2C) data ecosystems are either still in their seminal stages or not even considered. Our extensive literature analysis has shown that, up until now, there are hardly any papers published concerning the integration of PD in data ecosystems. Moreover, information systems research almost entirely lacks design knowledge in this regard (Koskinen et al., 2019). To that end, we define our research methodology to propose an initial set of applicable design principles for building B2C data ecosystems with data sovereignty, making them human centric. Following, we use the linguistic abbreviation *B2C data ecosystem*, not mentioning human centricity explicitly.

3. Research design

3.1 Design principle construction method

Design Principles (DPs) provide prescriptive guidance for action to design an artefact more efficiently (Möller et al., 2020). There are various approaches to develop DPs reflectively, e.g., by formalising and codifying experts' experiences (Azkan et al., 2021). In this study, supportive DPs are developed *a priori*. They intend to support the construction of human centric B2C data ecosystems by synthesising data from both the literature and the field.

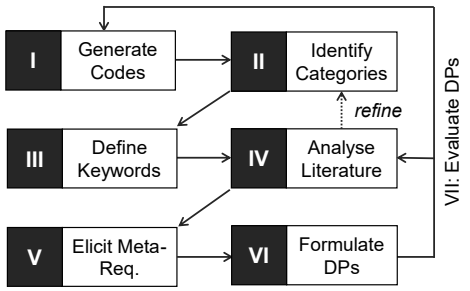

Figure 1. Research design in alignment with Möller et al. (2020).

Our study is structured based on Möller et al. (2020) who propose a method explicitly tailored to generate *supportive* DPs. The presented research is part of a broader Design Science Research (DSR) project that comprises several artefacts in the field of B2C data ecosystems (e.g., taxonomies, reference models, and instantiations). The orientation towards

DSR is reasonable since the methodology is suitable to solve problems of the real world by systematically designing relevant artefacts (Hevner et al., 2004). Our DSR approach is based on Kuechler and Vaishnavi (2008), consisting of the phases problem awareness building, solution suggestion, artefact development, and its evaluation, while encompassing multiple design cycles. This DSR methodology is embedded into the study's research design shown in Figure 1. In the following, its intermediate steps are outlined.

Our research design chosen to develop supportive DPs embodies two central research methods. Firstly, we conducted an interview study to narrow down our focus and the domains to be analysed. This is vital, since sharing and processing PD in general, and within B2C data ecosystems in particular, touches upon fundamental topics in economics, ethics, and law, thus exceeding mere technical considerations (Meister & Otto, 2019). Secondly, building upon the implications of the interviews, we carried out a structured literature review (SLR) with search strings composited by overarching code themes as keywords that were derived from the interview study. Literature analysis enabled a thematic deep dive into those themes, framed as *categories*, a further delimitation of their relevant aspects and, ultimately, an itemization and aggregation of their underlying concepts that allowed to formulate DPs (Möller et al., 2020).

3.2 Identification of knowledge base

Qualitative interviews with experts are an established method to generate data in the IS field (Schultze & Avital, 2011). We relied on qualitative data collected from interviewees with expertise in the domains data sovereignty and (B2C) data ecosystems who are advanced in at least one of the relevant dimensions addressed by Meister and Otto (2019), i.e., technology, economics, ethics, and law (see Table 1).

Since DPs address a design purpose, they need an addressee who uses them. This means-end relationship determined our selection of interviewees. As B2C data ecosystems represent a widely unexplored research area (Koskinen et al., 2019), interviews served to solve the rather academic question of "where to look into", while DPs themselves were extracted from literature (see Figure 1). Thus, interviewing human end users and involving them in the design process becomes effective in later stages of the aforementioned broader DSR project. In this study, we opted for experts of associated domains (Meister & Otto, 2019) while emphasizing scientists in our interviewee sampling due to the explorative nature of our research.

As expert sampling approach, we invited experts from our personal networks (Bhattacherjee, 2012). The interviews were mostly conducted in the year 2021. *A priori*, we prepared an interview guide to ensure that a similar range of topics is discussed with the interviewees, depending on their background, i.e., their dimension of expertise, making the interviews semi-structured (Merton & Kendall, 1946). From an abstract point of view, the interview guide comprised four main steps which were discussed sequentially. Firstly, experts were asked to describe human centric B2C data ecosystems from their point of view in detail. Secondly, they had to suppose possible implications arising for those systems in their domain of profession, e.g., data law, privacy-preserving technology, and data markets. Thirdly, experts were asked to propose further relevant domains and topics they associate with B2C data ecosystems and corresponding implications which were not necessarily in the experts' profession. Lastly, interviewees ranked the importance of stated implications for B2C data ecosystems and possible interrelations between them were discussed.

From the coding of interviews, we elicited overarching categories by means of the *Straussian* approach, i.e., the sequential entanglement of open, axial, and selective coding (Strauss & Corbin, 1990). Those thematically differentiated categories, arising from sorting codes to themes, allowed to initiate the SLR (see Figure 1). The SLR encompassed several precisely tailored search strings to extend and verify the categories and to derive meta-requirements.

Table 1. Interviewed experts by domain, role, discipline, dimension, and interview duration.

Domain	Position	Discipline	Relevant Dimension	[min]
Industry	CEO	Ethics, Management, Medicine	Ethics	48:55
Research	Project Manager	IT/IS Engineering	Economics, Technology	38:28
Industry	CEO	IT/IS Engineering, Management	Economics, Law, Technology	44:38
Research	Project Manager	Economics	Economics	36:07
Research	Professor	Economics, IT/IS Engineering	Economics, Technology	41:41
Research	Professor	Medicine, Ethics, Humanities	Ethics, Law	42:35

3.3 Elicitation of meta-requirements

Meta-requirements (MR) are necessities that apply to a class of artefacts rather than a single instance alone (Möller et al., 2020). By substantiating the findings of the interviews through descriptive evidence encountered in literature, we generated meta-requirements for B2C data ecosystems and, ultimately, formulated supportive DPs (Möller et al., 2020).

Following recommendations of Webster and Watson (2002), we conducted our SLR in *Scopus*, while the *AIS e-Library* was used for cross-checking results. Scopus and AIS are suitable, as they are large multidisciplinary databases covering published material, above all, in the humanities and social sciences. Furthermore, one can implement long and very precise search strings (Falagas et al., 2008), which is particularly relevant in our research design.

We used the terms directly derived from coded interview material, i.e., the categories, for our search strings. We searched for those keywords in "titles", "authors' keywords", and "abstracts" of periodicals. We analysed literature by extracting phrases with useful content that further extended, refined, or verified our categories. Those phrases were coded, included in a system of tables (in Microsoft Excel), and iteratively generalized. A code, in that regard, means the "*construct that symbolizes and thus attributes interpreted meaning to each datum for later purposes of pattern detection, categorization, theory building, and other analytic processes*" (Saldaña, 2021, p. 4), which makes it a suitable tool for our purposes. Problems occurring during phrase extractions (or coding in interviews) were discussed among authors directly until a consensus was reached.

We leveraged our findings to further narrow down the categories originally identified in the interviews, thus consistently extending the knowledge base. Building upon the concretized categories and their underlying concepts, we elicited meta-requirements and, ultimately, defined DPs. As shown in Figure 1, expert interviews and literature analyses were carried out in several iterations, leading to a continuous refinement of categories, meta-requirements, and DPs.

4. Design principles for human centric B2C data ecosystems

4.1 Meta-requirements

Our research outcome is dividable into a set of meta-requirements (MR) and subsequently developed supportive DPs. Möller et al. (2020) state that meta-requirements are a mandatory part of DP development endeavours as they ensure *value grounding*, meaning every DP is justified by at least one encountered meta-requirement that it aims to fulfil (Goldkuhl, 2004).

Meta-requirements must be generalized to a sufficient extend to become transferable to multiple artefacts of the same class. Likewise they are inevitably decoupled from their application scenarios once derived from (Walls et al., 1992). Table 2 summarizes our proposed meta-requirements in their most aggregated form, addressing significant challenges in building B2C data ecosystems. Our research revealed that, except for a few initial approaches like SOLID or the Data Transfer Project, and some publicly funded research projects, there are hardly any progresses towards developing B2C data ecosystems in practice yet. Additionally, literature widely lacks corresponding design knowledge.

Our DPs represent linguistic and prescriptive statements for action that respond to the elicited meta-requirements. Thus, DPs are derived from the MRs, which, in turn, stem from investigations in practice (i.e., the interviews) and theory (i.e., the SLR). To structure our set of DPs, we draw from the dimensions of humans' data sovereignty proposed by Meister and

Table 2. Meta-requirements by domain with definition.

#	Themes	Description of Meta-Requirement
MR1	Data Sovereignty	*In a B2C data ecosystem, data subjects must control all sharing and processing activities concerning their data and receive adequate compensation thereof.*
MR2	Legal Compliance	*A B2C data ecosystem must comprise (quasi-) legal instructions and restrictions for the enforcement of the data subjects' rights associated with their data, whereby the specificities of those legal instructions depend on the applicable jurisdiction.*
MR3	Economic Rationale	*A B2C data ecosystem must enable data consumers to leverage data as corporate assets within the system, encompassing their systematic generation, collection, trading, analysis, processing, and linking.*
MR4	Ethical Correctness	*A B2C data ecosystem must constantly search for and evaluate possible moral problems related to PD processing within the system and provide (binding) guidelines for all actors describing how they have to handle data ethically, e.g., a Code of Conduct.*
MR5	Technical Implementation	*A B2C data ecosystem must provide a technically feasible environment for systematically sharing, monetizing and utilizing PD that is secure and actively integrates the human users while exhibiting a high degree of usability for all actors.*

Otto (2019) and Lauf et al. (2022), i.e., technology, economics, law, and ethics. Using these dimensions ensures that the DPs cover, comprehensively, issues determining the design of B2C data ecosystems. We rely on the authors' higher-level meta-dimensions since we consider humans' data sovereignty as a sine qua non in the design of B2C data ecosystems (see Section 2). Based on the preceding MRs and the guiding concept of data sovereignty, we derived seven supportive DPs that are explicated in the following.

4.2 Supportive design principles

DP1: *Provide a mechanism enabling data subjects to shift as many tasks as legally possible to a "deputy-actor" in the ecosystem, e.g., a data trust, to maximize usability through process* ***automation****, while the scope of action attributable to this deputy depends on the applicable jurisdiction, i.e., European data law.*

In B2C data ecosystems, systematic sharing and utilization of PD inevitably lead to the data subjects being confronted with a myriad of tasks requiring them to extensively engage with the system. One can assume that the number of emerging tasks will far exceed their processing capacities and digital competencies (Bester et al., 2016). Thus, the provision of usability arises as a critical issue, making mechanisms for automation crucial. Since automation in the context of B2C data ecosystems inevitably implies a *broad consent* of data subjects, a legally applicable solution is of utmost importance. Taking European data law as our guiding example for a restrictive legal framework, consent can only be formulated broadly for situations not requiring a *specific consent* of the individual. This entails a hybrid nature of the underlying consent model whose design characteristics will always be subject to the respective jurisdiction. Design decisions for this model should be taken under the aegis of legal experts and, eventually, ethicists, depending on the sensitivity of the data. First suggestions are provided by, for instance, Geller et al. (2022), albeit limited to the clinical context.

DP2: *Provide a mechanism for* ***incentivization*** *of data sharing to ensure economic viability of the ecosystem and promoting joint data utilization.*

To facilitate sharing and utilization of PD, a B2C data ecosystem requires an appropriate incentivization mechanism. On the one hand, this mechanism must encourage humans to share data while, on the other hand, providing insights into the data for data consumers without curtailing privacy. In principle, the applied incentivization must ensure data subjects a "fair" compensation for (temporarily) disclosing data within the ecosystem, given an acceptable purpose of an inquiring data consumer. The prevailing problem in that regard is the determination of an objectively (or approximately) fair value of PD to be offered to the data subject. A naive solution is to disclose the metadata, particularly measures of data quality, e.g., by means of a catalogue system. Subsequently, data consumers can offer a price based on their individual Willingness-to-Buy. Under the assumption that data consumers can define such a valuation based on their expected value creation through data processing, this solution must always undergo an examination with regard to whether the data sovereignty of the individuals in the B2C data ecosystem might be restricted. A problem arising in this context is their lack of digital literacy as humans are hardly aware of the economic value linked to their data (Section 1). Thus, an incentivization mechanism must support transparency by facilitating information for all actors, ideally encompassing data quality computations by (trusted) third parties. For such computations, *IBM Cloud Pak for Data* and *IBM Information Analyzer* are vivid business examples of how such a third party service in a B2C data ecosystem might look like.

DP3: *Provide a technically enforceable* ***data governance*** *structure by integrating an effective usage control framework that enables data subjects to sovereignly control data processing in the ecosystem.*

MR1 and MR2 prescribe that data subjects must always be in charge of all processing activities related to their data. Thus, access control is not expedient in B2C data ecosystems. As a more severe control option, usage control must be enforced, which inevitably comprises the restriction that PD must not pass system boundaries. Consequently, irreversible anonymization arises as the only possibility for data to leave the ecosystem. Relying on pure legal agreements for usage control, instead of a technical enforcement, is not suitable in a B2C context. This is because data subjects would need both the digital literacy and the willingness to constantly trace a myriad of datasets shared with data consumers and check conducted processing activities, respectively purposes (e.g., Art. 30 GDPR), against mutually agreed usage terms. Thus, usage control must be technically enforced implying impermeable ecosystem boundaries.

DP4: *Leverage a scalable and* ***decentralized infrastructure*** *with data stored in the sphere of the data subjects to increase their control of data as well as trust in and (data) security of the system.*

B2C data ecosystems should not rely on a monolithic centralized infrastructure but rather leverage decentralized infrastructures that are open, distributed and shared. Likewise, the infrastructure

may function as a common and be governed in a democratic and self-determined manner (Nagel et al., 2021). Moreover, it should be built upon agreements between actors and, by its design, favor data security, privacy and trustworthiness of the system. Relying on decentralized infrastructures counteracts the current mode of operation with respect to PD handling, which is characterized by a limited number of centralized providers and the concentration of data in a few hands (Lauf et al., 2022; Nagel et al., 2021). A scalable and decentralized infrastructure supports a level playing field for sovereignly sharing and utilizing PD within a human centric data ecosystem. Best practices such as SOLID, IDS and Gaia-X should be used as orientation.

DP5: *Provide a mechanism for **standardized data processing** to utilize data securely and in a privacy-preserving way within the system while allowing consumers to analyse data with respect to their needs.*

In a B2C data ecosystem, data consumers must be able to process data while data usage restrictions are in place and technically enforced (DP3). This is crucial to impede processing activities exceeding purposes the data subject and the consumer have agreed upon at the time of data sharing. Such processing restrictions add additional complexity to the system's data governance framework. The reason is that technically controlled data processing is only realizable by means of data apps executed solely within system boundaries. Different to B2B settings, in B2C ambits, those apps should be standardized in the sense that they are directly assignable to specific usage purposes (i.e., restrictions) data subjects can define. This, in turn, implies that in B2C data ecosystems data subjects must be limited by their possibilities to specify usage policies. Otherwise, the heterogeneity of usage restrictions attached to various datasets precludes the majority and the standardization of data processing activities. Data apps are an already applied technique to implement largely standardized processing operations in data ecosystems, e.g., Gaia-X and IDS. However, in B2C ambits, they must be irrevocably bound to specifiable usage restrictions (appropriation), for which there are no examples in practice yet.

DP6: *Provide a mechanism that supports ethical compliance of personal data processing in the system by defining and enforcing behavioral guidelines (i.e., a **Code of Conduct**) all actors must obey.*

Since processing PD should always have an ethical grounding, basic moral guidelines must be formulated and obeyed in B2C data ecosystems. Thus, equal treatment of actors (e.g., no discrimination based on data) and responsible data processing (e.g., evaluation of processing purposes for their adherence to societal values) must be ensured rigorously. Attention should be drawn to establish transparent communication as well as decision and escalation lines in governance to support trust of all actors while making them feel appreciated (Nagel et al., 2021). Desired ways of treating both data and actors in the B2C data ecosystem should be included in a *Code of Conduct* all actors must accept before being admitted to the system, e.g., during an onboarding procedure. Adherence to that code should be objectively observable in the system, for instance, by means of repositories of data processing purposes (e.g., Art. 30 GDPR) and recorded consent proofs.

DP7: *Provide a mechanism ensuring semantic **uniformity** of actors, services, and data in the ecosystem to support interoperability, findability, processability, and portability of data and services as well as to facilitate communication among actors.*

Due to their decentral nature (DP4), B2C data ecosystems require entities that orchestrate vocabularies (i.e., ontologies, reference data models, or metadata) applicable to annotate and describe datasets, services and, eventually, actors themselves. Thus, an information model is needed as the basis for, firstly, the description of data sources and, secondly, the provision of other domain specific vocabularies. In this context, using metadata is advisable to support interoperability, findability and portability of datasets in the system. According to common practices in B2B contexts (Otto et al., 2019), B2C data ecosystems should reuse existing domain vocabularies and standards where possible to foster acceptance of actors and interoperability of data and services. Furthermore, as data exchange between different actors is at the core of B2C data ecosystems, only a fundamental set of vocabularies for data descriptions and data exchange invocations might be required. Additional domain-specific vocabularies should be provided wherever necessary to extend those concepts and to offer more information about data provided or requested. Ultimately, different actors relying on the same vocabularies to structure and describe data and services circulating in the B2C data ecosystem foster its semantic consistency and lead to a significant increase of functionality and operativeness.

5. Interplay of meta-requirements and design principles

The supportive DPs are directly connected to the derived meta-requirements as described in this section. The MR **Data Sovereignty** contributed content to three DPs that address both the protective claims (i.e., DP3, DP4) and the participatory claims (i.e., DP2) of

data subjects to data. Furthermore, the MR **Legal Compliance** affects almost the entire set of generated DPs, accentuating the central importance of data protection and privacy in B2C data ecosystems. The consideration of the applicable jurisdiction, e.g., European data law, in almost all DPs (i.e., DP1, DP3–DP6) distinguishes B2C from B2B ambits. In fact, the latter commonly tries to avoid any involvement of PD and the legal obstacles arising in their peripheries.

Since existing concepts for systematically sharing and utilizing PD, e.g. personal data markets, exhibit economic viability issues (Spiekermann et al., 2015), particular attention must be drawn to the **Economic Rationale** of B2C data ecosystems. The homonymous MR results in DPs addressing aspects of data economics (i.e., DP2) and data sovereignty (i.e., DP5), whereas the latter addresses the limitation of possible processing purposes. Furthermore, sharing and utilizing PD within economic structures requires an ethical and a philosophical evaluation (Spiekermann et al., 2015). Thus, **Ethical Correctness** arises as a MR that is difficult both to implement and to assess in B2C data ecosystems. Even though the evaluation of moral issues lies in the eye of the beholder, we suggest a *Code of Conduct* as related DP. In combination with the rigor orientation towards data sovereignty (MR1) and effective data governance structures (DP3), this DP supports ethical correctness of the admittedly critical system under consideration.

Technical Implementation encompasses not only infrastructure decentralization of the ecosystem (i.e., DP4) with its organizational (i.e., DP5) and semantic (i.e., DP7) conditions but also accentuates the decisive criticality of system usability for human users (i.e., DP1). Accordingly, B2C data ecosystems must find user-friendly solutions while ensuring data sovereignty, legal compliance, and economic viability. This is difficult, since B2C data ecosystems are characterized by data subjects inevitably facing a myriad of tasks that will exceed their processing capacities and, frequently, their digital competencies (Bester et al., 2016). A conceivable solution is process automation (DP1) by integrating a service provider as deputy actor, e.g., a data trust, for taking over (parts of) the data subjects' obligations. The implementation of such a proxy solution is likely to become legally difficult, depending on the restrictiveness of the underlying jurisdiction. Though, even under restrictive European data law, as our exemplary legal lens for examining B2C data ecosystems, applicable solutions for such a proxy have already been conceptualized by former research. Blankertz (2020), Funke (2020), and Reed et al. (2019) propose data trust models as neutral and trusted service providers tailored to European data law. Such models are applicable to orchestrate PD for the data subjects in a data ecosystem. However, there are no examples in practice for the integration of data trusts in data ecosystems yet. To the best of our knowledge, data trust models in ecosystem contexts are currently not explicitly regulated under European data law and only conceivable as a hybrid solution. We consider a hybrid solution as expedient whenever *specific consent* of the data subject is required in any process of the ecosystem ((Funke, 2020); DP1).

6. Discussion and conclusion

Since by now the concept of data ecosystems has emphasized the B2B context, IS research almost entirely lacks design knowledge for the integration of humans in such systems. Likewise, there are only very few initial developments of B2C data ecosystems in practice which are far from marketability. This results in an impairment of growth and innovation potentials of the information economy. We address the issue and conceptualize initial design knowledge into five MRs and seven DPs supporting practitioners and scientists in understanding and successfully conceptualizing B2C data ecosystems. By doing so, existing research is taken a step further and insights into this hardly explored and interdisciplinary field are offered.

The **practical contribution** of the study is to provide basic information on key aspects and principles of how B2C data ecosystems should be designed from a high level of abstraction. These help to reflect on own approaches and ideated concepts on the one hand and, on the other hand, to design B2C data ecosystems from scratch. The DPs show which aspects have to be considered when developing B2C data ecosystems. In particular, they highlight the importance of human centricity (i.e., data sovereignty) as the guiding concept while, among others, providing possibilities to shift obligations of data subjects to legally compliant proxy solutions. By ensuring usability of the complex decentral system through such mechanisms, individuals are enabled to actually participate in data ecosystems. Albeit the DPs are held generic to offer flexibility, they offer concrete recommendations of action due to the almost entirely unexplored research area they address. For **scientific contributions**, the description of the design process enables a scientific validation and extension of the artefact. The DPs are the first step into generating comprehensive design guidelines for human centric B2C data ecosystems and serve as a foundation stone for further work. They produce inductive insights and deep understanding of implications arising when integrating data sovereign humans in data ecosystems.

The study is subject to the following **limitations**. As the research area encompasses an interdisciplinary

and wide field of topics, we must assume to have analyzed a fraction of associable literature only. Thus, the collection of (meta) requirements might not be conclusive. Additionally, we emphasized European data law as our exemplary legal lens and neglected implications given by other, typically less restrictive, jurisdictions. Consequently, much higher potentials for B2C data ecosystems are likely to be encountered in non-European legal frameworks due to our focus on an extreme case. Moreover, as meta- requirements and DPs were largely conceptualized by means of desk research, mainly published material was used. This inevitably means that the results could only be built on what is publicly available. Furthermore, the qualitative nature of our study entails subjectivity issues in terms of the coding of interviews, the derivation of meta-requirements, and the formulation of DPs. In summary, the limitations suggest that our DPs might not yet entirely encompass the very broad and interdisciplinary research area of B2C data ecosystems but rather represent a first approach to create design knowledge in this hardly explored field. Notably, we do not consider our lens on European data law as a constraint for the meaningfulness of our research since the restrictive nature of this jurisdiction makes both our DPs and our MRs applicable in less restrictive legal settings as well. Nevertheless, our results require further extension and verification as outlined in the recommendations for future research.

Future research should continue the subsequent design science research path by developing models and prototypical instantiations as well as conducting field tests to evaluate and refine the compiled DPs. By now, the results represent initial hypotheses that need to be validated. Additionally, more data sources (e.g., further interviews, case studies and literature analyses) should be used in additional domains, to triangulate a more comprehensive look into B2C data ecosystems.

The results of our empirical study are promising and provide substantial knowledge for improvements in following design cycles. It guides practitioners and scientists in developing and theorizing human centric B2C data ecosystems while supporting scientific comprehension of the design processes.

References

Abu-Matar, M. (2016). Towards a software defined reference architecture for smart city ecosystems. *2016 IEEE International Smart Cities Conference*, 1–6.

Adonis, A. A. (2019). Critical Engagement on Digital Sovereignty in International Relations: Actor Transformation and Global Hierarchy. *Global: Jurnal Politik Internasional*, *21*(2), 262–282. https://doi.org/10.7454/global.v21i2.412

Azkan, C., Iggena, L., Möller, F., & Otto, B. (2021). Towards Design Principles for Data-Driven Services in Industrial Environments. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 1789–1798.

Bester, J., Cole, C. M., & Kodish, E. (2016). The limits of informed consent for an overwhelmed patient: clinicians' role in protecting patients and preventing overwhelm. *AMA Journal of Ethics*, *18*(9), 869–886.

Bhattacherjee, A. (2012). *Social science research. Principles, methods, and practices*. Scholar Commons, University of South Florida; Open Textbook Library.

Birch, K., Cochrane, D. T., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society*, *8*(1).

Blankertz, A. (2020). *Designing Data Trusts. Why We Need to Test Consumer Data Trusts Now*. https://www.stiftung-nv.de/sites/default/files/designing_data_Trusts_e.pdf

Couture, S., & Toupin, S. (2019). What does the notion of "sovereignty" mean when referring to the digital? *New Media & Society*(21), Article 10, 2305–2322.

Currie, W. S. (2011). Units of nature or processes across scales? The ecosystem concept at age 75. *New Phytologist*, *190*(1), 21–34.

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.

Fezer, K. H. (2017a). Dateneigentum. *Multi Media Und Recht*(1), 3–5.

Fezer, K.-H. (2017b). Dateneigentum der Bürger. *Zeitschrift Für Datenschutz*(3), 99–105.

Funke, M. (2020). *Die Vereinbarkeit von Data Trusts mit der Datenschutzgrundverordnung (DSGVO)*. https://algorithmwatch.org/data-trusts/

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. *BIOSTEC 2022*, *5*, 81–92.

Goldkuhl, G. (2004). Design Theories in Information Systems - A Need for Multi-Grounding. *Journal of Information Technology Theory and Application*, *6*(2), 59–72.

Hanssen, G. K., & Dyba, T. (2012). Theoretical Foundations of Software Ecosystems. *Proceedings of the International Workshop on Software Ecosystems*, 6–17.

Hevner, March, Park, & Ram (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hummel, P., Braun, M., Augsberg, S., Ulmenstein, U. V., & Dabrock, P. (2021). *Datensouveränität: Governance-Ansätze für den Gesundheitsbereich* (1st ed.). Springer Nature.

Hummel, P., Braun, M., & Dabrock, P. (2021). Own data? Ethical reflections on data ownership. *Philosophy & Technology*, *34*(3), 545–572.

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, *39*(8), 2255–2276.

Jansen, S., Cusumano, M. A., & Brinkkemper, S. (2013). *Software ecosystems: analyzing and managing business networks in the software industry*. Edward Elgar Publishing.

Koskinen, J., Knaapi-Junnila, S., & Rantanen, M. M. (2019). What if we had fair, people-centred data economy ecosystems? *SmartWorld / SCALCOM / UIC / ATC / CBDCom / IOP / SCI*, 329–334.

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489–504.

Lauf, F., Scheider, S., Bartsch, J., Herrmann, P., Radic, M., Rebbert, M., Nemat, A. T., Schlueter-Langdon, C., Konrad, R., Sunyaev, A., & Meister, S. (2022). Linking Data Sovereignty and Data Economy: Arising Areas of Tension. *International Conference on Wirtschaftsinformatik*, *17*.

Lee, D. (2014). Building an open data ecosystem: an Irish experience. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 351–360.

Leidner, D. E., & Tona, O. (2021). The CARE Theory of Dignity Amid Personal Data Digitalization. *MIS Quarterly*, *45*(1), 343–370.

Li, W., Badr, Y., & Biennier, F. (2012). Digital ecosystems: challenges and prospects. *Proceedings of the International Conference on Management of Emergent Digital Ecosystems*, 117–122.

Lindman, J., Kinnari, T., & Rossi, M. (2015). Business roles in the emerging open-data ecosystem. *IEEE Software*, *33*(5), 54–59.

Lohsse, S., Schulze, R., & Staudenmayer, D. (Eds.). (2020). *Data as Counter-Performance – Contract Law 2.0? A Market Model for Personal Data: State of Play Under the New Directive on Digital Content and Digital Services*. Münster Colloquia on EU Law and the Digital Economy V.

Maritain, J. (1950). The concept of sovereignty. *The American Political Science Review*, *44*(2), 343–357.

Meister, S., & Otto, B. (2019). Digital life journey: A Framework for a self-determined life of citizens in an increasingly digitized world. *Basic Research Paper*. http://publica.fraunhofer.de/documents/N-559377.html

Merton, R. K., & Kendall, P. L. (1946). The Focused Interview. *American Journal of Sociology*, *51*(6), 541–557.

Möller, F., Guggenberger, T. M., & Otto, B. (2020). Towards a Method for Design Principle Development in Information Systems. *International Conference on Design Science Research in Information Systems and Technology*, 208–220.

Nagel, L., Lycklama, D., & Ahle, U. (2021). *Design Principles for Data Spaces* [Position Paper]. https://design-principles-for-data-spaces.org/

Oehler, A. (2016). Chancen der selbstbestimmten Datennutzung. *Wirtschaftsdienst*, *96*(11), 830–832.

Oliveira, M., Lima, G., & Lóscio, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. *Knowledge and Information Systems*, *61*(2), 589–630.

Otto, B., Steinbuß, S., Teuscher, A., Lohmann, & Steffen et al. (2019). *IDS Reference Architecture Model Version 3.0*. https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf

Pohle, J., & Thiel, T. (2020). Digital sovereignty. Internet Policy Review, *9*(4).

Rantanen, M. M., & Koskinen, J. (2020). Respecting the individuals of data economy ecosystems. *International Conference on Well-Being in the Information Society*, 185–196.

Reed, C., Solicitors, B. P. E., & Masons, P. (2019). *Data trusts: legal and governance considerations*. https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf

Saldaña, J. (2021). *The coding manual for qualitative researchers*. SAGE Publications Limited.

Sambra, A. V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., Zagidulin, D., Aboulnaga, A., & Berners-Lee, T. (2016). Solid: A platform for decentralized social applications based on linked data. *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep*.

Schmitt, C. (2005). *Political theology: Four chapters on the concept of sovereignty*. University of Chicago Press.

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. *Information and Organization*, *21*(1), 1–16.

Schwartz, P. M. (2004). Property, privacy, and personal data. *Harvard Law Review*, *117*(7), 2055–2128.

Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, *25*(2), 161–167.

Strauss, A., & Corbin, J. M. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. SAGE.

Véliz, C. (2020). *Privacy is power*. Bantam Press.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, *3*(1), 36–59.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), 13–23.

Werthner, H., & van Harmelen, F. (Eds.). (2017). *Informatics in the Future*. *Digital sovereignty and IT-security for a prosperous society*. Springer.