Kranz, Johann et al.

Data Portability

Business & Information Systems Engineering

Provided in Cooperation with:
Springer Nature

Suggested Citation: Kranz, Johann et al. (2023) : Data Portability, Business & Information Systems Engineering, ISSN 1867-0202, Springer Fachmedien Wiesbaden GmbH, Wiesbaden, Vol. 65, Iss. 5, pp. 597-607,
https://doi.org/10.1007/s12599-023-00815-w

Bus Inf Syst Eng 65(5):597–607 (2023)
https://doi.org/10.1007/s12599-023-00815-w

CATCHWORD

Data Portability

Johann Kranz · Sophie Kuebler-Wachendorff · Emmanuel Syrmoudis · Jens Grossklags · Stefan Mager · Robert Luzsa · Susanne Mayr

Received: 5 May 2022 / Accepted: 13 March 2023 / Published online: 22 May 2023

Keywords Data portability · Data transfer · Interoperability · Digital markets · Privacy

1 Introduction

Many markets for online services such as social media, search, social messaging, or commerce have developed into skewed playing fields, where a small number of dominant online service providers (OSPs) have gained enormous economic and epistemic power (Zuboff 2019). OSPs owning vast amounts of user data benefit from self-reinforcing advantages arising due to (data) network and lock-in effects (Gregory et al. 2021) that eventually make them "data monopolists" and virtually incontestable gatekeepers (Autor et al. 2020). The process is self-reinforcing because dominant OSPs can exploit the exponentially increasing amount of user data to create data-driven innovation and powerful lock-in effects. The ability to apply data-driven learning and advanced artificial intelligence methods enables dominant OSPs owning large proprietary data silos to continuously improve, innovate, and adapt their service offerings (Gregory et al. 2021). Thus, dominant OSPs' ability to meet and shape user demands continuously increases, while the ability of smaller rival OSPs to compete in the market, including those with services that are more respectful of users' privacy, continuously deteriorates.

As such, even if dominant OSPs unfairly exploit their market position or disregard user privacy and agency, the economic rationale for users to move to alternative online services is low due to high lock-in effects and switching costs (Easley et al. 2018; Sunyaev et al. 2021; Wohlfarth 2019). One such privacy challenge is the (re)use of data by OSPs for purposes deemed problematic or invasive by users, which also raises questions regarding data ownership and corresponding accountabilities (Fadler and Legner 2022). Despite ever-increasing high-profile privacy misconduct (e.g., revelations about Facebook's privacy practices by former employee Francis Haugen[1]) users are left with little meaningful options to adopt data protection and privacy measures and to move to rival OSPs due to the skewed playing field and high switching barriers.

Accepted after two revisions by Christine Legner.

J. Kranz (✉) · S. Kuebler-Wachendorff · S. Mager
Ludwig-Maximilians-Universität München, Munich, Germany
e-mail: kranz@lmu.de

S. Kuebler-Wachendorff
e-mail: kuebler-wachendorff@lmu.de

S. Mager
e-mail: stefan.mager@lmu.de

E. Syrmoudis · J. Grossklags
Technische Universität München, Munich, Germany
e-mail: emmanuel.syrmoudis@tum.de

J. Grossklags
e-mail: jens.grossklags@tum.de

R. Luzsa · S. Mayr
Universität Passau, Passau, Germany
e-mail: Robert.Luzsa@uni-passau.de

S. Mayr
e-mail: Susanne.Mayr@uni-passau.de

[1] See: https://www.commerce.senate.gov/services/files/FC8A558E-824E-4914-BEDB-3A7B1190BD49.

2 Portability Regulation

The *Right to Data Portability* (RtDP) stipulated in Art. 20 by the European Union in 2018 as part of the General Data Protection Regulation (GDPR) aims to level the playing field in digital markets.[2] Compared to already established rights such as access and correction in the previous Data Protection Directive from 1995, the RtDP was a major update of data regulation (Wong and Henderson 2019). Data portability aims to enable users to easily and securely transfer personal data from one service to another service and reuse it without any restrictions, and thereby advances users' opportunities to own, control, and manage their personal data (De Hert et al. 2018; Sunyaev et al. 2021). While the notion of personal data in the GDPR is broad and includes *"any information relating to an identified or identifiable natural person"* (Art. 4.1 GDPR), the RtDP is limited to personal information "*provided*" by the user to an online service. Art. 20 GDPR introduces two notions of data portability. First, users may "*receive*" their data following a RtDP request in a structured, commonly used, and machine-readable format, which would allow them to upload (parts of) the data to another service. Second, "*where technically feasible*", personal data should be directly transmitted to another service upon request by the user. These practically relevant aspects are discussed in more detail in the next section.

More generally, the RtDP aims at reducing the power of data monopolists and increasing competition and innovativeness in data-driven digital markets. As users can transfer their data to alternative OSPs, switching costs and lock-in effects should decrease. This should strengthen the competitiveness of smaller OSPs, since improved access to historical user data allows to generate more data-based value (Sunyaev et al. 2021; Wohlfarth 2019). Further, the RtDP enhances user choice and privacy, which can be defined as an individual's control over the acquisition and use of their personal information (Pavlou 2011). However, it must be noted that the RtDP can only achieve the desired effect on data and privacy protection in combination with other rights of data subjects under the GDPR, such as the right to erasure to avoid the spread of their data over multiple OSPs (Rupp et al. 2022).

The RtDP can be seen as part of the broader EU Data Strategy that aims at creating a thriving single data market within the EU and across sectors to increase data-driven innovativeness and competitiveness. Respective legislative proposals that will soon come into effect include the Digital Markets Act, Digital Services Act, Data Act, Data Governance Act, and Artificial Intelligence Act. For data portability, especially the Data Act and Digital Markets Act are relevant as they broaden the initial scope of the RtDP.[3] The Digital Markets Act embodies an asymmetric regulation approach that poses strict requirements for dominant OPSs referred to as "gatekeepers" while exempting smaller rivals. Gatekeepers are defined as providers of a core platform service that acts as a significant gateway for businesses to reach end users and benefits from an entrenched and durable market position. The Digital Markets Act mandates that gatekeepers provide effective tools to facilitate data portability, including real-time, high-quality, and continuous access to data generated by engaging with gatekeepers' services and products. The Data Act additionally broadens the scope of data portability as not only personal data is included, but also datasets including a mix of personal and non-personal data generated by objects connected to the Internet of Things. This includes data obtained, generated, and collected by networked objects such as vehicles, home equipment and consumer goods, medical and health devices, or agricultural and industrial machinery regarding their performance, use, or environment. However, devices that are primarily designed to record and transmit content such as personal computers, servers, smart phones, cameras, and sound recording systems should not be covered by this regulation. Hence, the Data Act focuses on enabling users and OSPs to receive access and extract value from data provided by the Internet of Things.

As a result of the new regulatory framework, the RtDP is reinforced, particularly for OSPs, and becomes significantly broader. The new regulation empowers users to switch services or multihome more easily which should lead to more innovation, competition, and choice in digital markets. Gatekeepers on the other hand need to invest in new technical solutions to comply with the new mandates.

Given this background, we aim at explaining the inherent promises and emerging multi-level challenges related to the RtDP. We believe that the topic offers rich research opportunities for the BISE/IS community to create impact by developing strategies for enhanced transparency, innovativeness, and competition in digital markets. Furthermore, data portability can serve as a guiding principle for meaningful consumer protection and data governance concepts that strike a balance between user privacy and innovation to increase "data richness" as envisioned by advocates of data sovereignty (Jarke et al. 2019).

[2] Similar regulations have been adopted in California with its California Consumer Privacy Act, in China in Article 45 of their Personal Information Protection Law, and in India and Brazil within their Personal Data Protection Bill and Lei Geral de Proteção de Dados Pessoais, respectively.

[3] See: DMA (COM 2020/842/EU, Art. 6.1 (h)) and DA (COM 2022/68/EU, Art. 5–7).

Table 1 Categorization of personal data (based on De Hert et al. 2018)

Data category	Covered by the RtDP	Data type	Description	Example
Provided personal data by user	Yes (narrow)	Received	Direct inputs by users	Search for a pizzeria nearby
	Yes (broad)	Observed	Collected by sensors	GPS coordinates, timestamp
Derived personal data by OSP	No	Inferred	Created by the OSP based on controlled data	User preferences (diet, time, area, budget)
	No	Predicted	Anticipates future prospects	Predictions of future user preferences (change in diet and budget with age)

3 Status Quo

Several studies about the status quo of the RtDP's implementation have unraveled three main obstacles: lack of user awareness and motivation, OPSs' reluctance to implement advanced import solutions, and a lack of standardization (Kuebler-Wachendorff et al. 2021). We will elaborate on the fundamental concepts of data portability and the status quo regarding implementation and adoption.

3.1 Data Scope

The GDPR mandates that OSPs have to export a user's *"personal data concerning him or her, which he or she has provided"*. Hence, the GDPR does not explicitly stipulate the concrete scope of personal data included in the RtDP (see Table 1). In a narrow sense, data portability only incorporates data that OSPs receive actively from users (e.g., address, bank account number), whereas in a broader sense it additionally includes observed data such as location data. However, inferred and predicted data derived from received and observed personal data is not covered by the RtDP (De Hert et al. 2018; Krämer 2020) or Data Act, although other rights stipulated in the GDPR like the "right of access" (Article 15) or the "right to erasure" (Article 17) include inferred data of users (European Data Protection Board 2022). The exclusion of data relating to inferences and predictions about the user limits the effectiveness of the RtDP, but maintains incentives for data-driven innovation of OSPs as such data remains protected (Engels 2016).

A recent analysis of the scope of data transferred by OSPs in response to a portability request found that for services that provided a compliant data export, 36% only contained received data, 55% additionally contained observed data, and 9% even contained inferred data (Syrmoudis et al. 2021). Further, the study indicates that the export scope of dominant OSPs is significantly higher than the export scopes of smaller rivals (Syrmoudis et al. 2021). This empirical finding is surprising since dominant OSPs are suggested to be negatively affected by data portability because rivals and new entrants can use data portability to attract new users and gain access to user data (Wohlfarth 2019). As such, the incentives of dominant OSPs to comply with the RtDP should be limited, particularly given the RtDP's lack of regulatory control and sanctions.[4]

3.2 Implementation

The majority of OSPs do not comply with GDPR's portability regulations, let alone that (smaller) OSPs regard the RtDP as a means for attracting users from (larger) rivals (Syrmoudis et al. 2021). The low level of utilization of the RtDP is likely partially driven by the lack of precision in the legal text of the GDPR. In particular, an important aspect to highlight is that the RtDP ought to comprise two approaches: direct and indirect data portability as illustrated in Fig. 1.

Direct data portability grants users the "*right to have the personal data transmitted directly from one controller to another"* (Art. 20 (2) GDPR), "*where technically feasible*". In contrast, indirect data portability is specified by the user's right to "*receive personal data concerning him or her"* (Art. 20 (1) GDPR) which users have to subsequently import manually at other OSPs. The Digital Markets Act is more specific as it requires real-time, high-quality, and continuous access to data which can only be achieved by direct data portability.

It is important to note that direct data portability, as intended by the regulation, requires currently nonexistent solutions that directly connect services of different OPSs. Consequently, users that want to make full use of the RtDP today need to transfer their data indirectly, as a direct, automated export and import from one service to another is not yet feasible (Syrmoudis et al. 2021). Yet, the difficult and time-consuming task of using the RtDP in an indirect fashion acts as an important complication for users as less than a third of OSPs comply with GDPR's export

[4] As of this writing, the "CMS. Law GDPR Enforcement Tracker" lists only two imposed penalties for noncompliance with GDPR Article 20, and both penalties related to violations of several GDPR Articles (see https://www.enforcementtracker.com/).

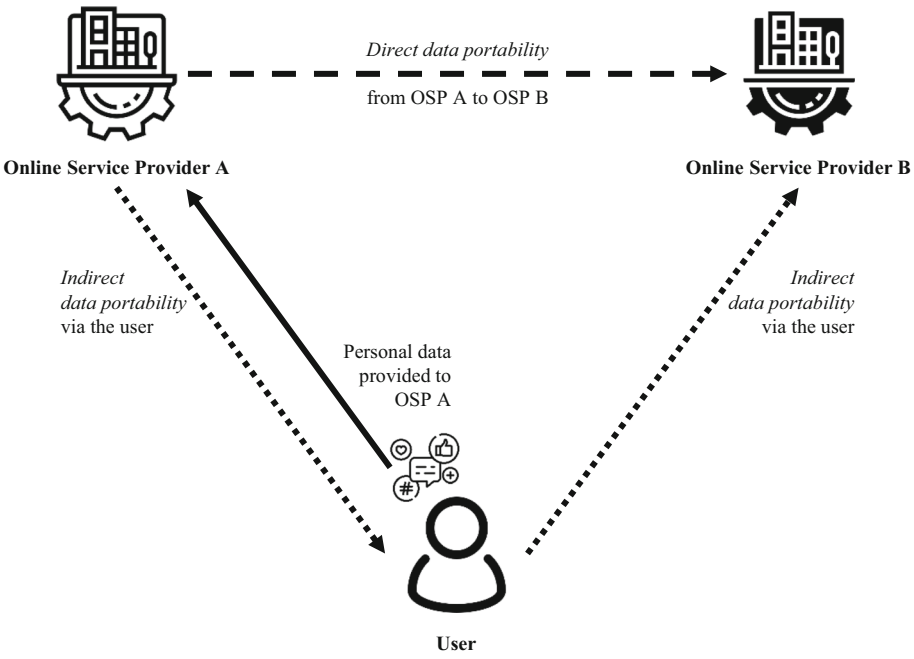

Fig. 1 Direct and indirect data portability

requirements and 76.8% of OSPs do not offer any import possibilities (Syrmoudis et al. 2021).

Further, to be compliant with Art. 20, OSPs have to export data in a "*structured, commonly used and machine-readable format*".[5] For OSPs that seek to import data, the usage of compliant file formats is crucial for allowing an automated or semi-automated processing of data. For some types of data, specific file formats have been standardized, like ICS for calendars or VCF for contacts. The use of standardized single-purpose formats allows OSPs to import data without needing to develop import procedures for each OSP exporting user data.

For compliant general-purpose formats such as JSON or XML, tools exist for parsing and transforming data (e.g., XSLT) and for standardization (e.g., DTD and XML Schema). Especially for XML Schema, standard formats for a wide range of purposes have been defined. Art. 20 does not mandate the use of (specific) standard formats. For example, a compliant data export could therefore be an XML or JSON file using a scheme that is unique for the exporting OSP. Importing OSPs would then need to develop separate import procedures for each exporting OSP they want to support. However, to date, OSPs do not provide public documentation about how they export data (Syrmoudis et al. 2021). Therefore, OSPs who want to offer import of data from other OSPs have to request an export themselves to learn how the exporting OSP currently structures its data exports. But, if an exporting OSP changes the structure of data exports without prior notice, data import will not function and importing OSPs are forced to redevelop import procedures.

3.3 User Adoption

From a users' perspective, we know that users' awareness and motivation to use privacy self-management systems and rights is often low – even if privacy concerns are high (Acquisti et al. 2020; Pavlou 2011). Correspondingly, less than a third of the respondents of a recent survey indicated that they were aware of the RtDP (Luzsa et al. 2022b). Asked about their ability to switch OSPs, about 25% reported that although they intend to switch OSPs because of general trust, privacy, and security concerns and to transfer their data to a new service, they actually failed to do so. The main barriers to switching OSPs were concerns about loss of social contacts, data, and content, as well as little knowledge about alternative OSPs or lack of experience with service switching–all of which proper implementations of the RtDP may help to alleviate. Moreover, further research revealed correlations between user characteristics and users' perceptions of the RtDP (Luzsa et al. 2022a). Users who describe themselves as very interested in and capable of using digital technology and to whom privacy is very important are particularly interested in making use of the RtDP to transfer their data between OSPs. Conversely, less technologically competent and less privacy-aware users tend to be hesitant towards the concept of data portability. In sum, current research on the user-side

[5] Compliant file formats are, in particular, XML, JSON, CSV, EML, ICS, MBOX, TEX, and VCS (Wong and Henderson 2019). Non-compliant or ambiguous formats include DOC/DOCX, PDF, and PNG due to their lack of structure and machine-readability.

of data portability suggests that the RtDP is still rather unknown and mostly appeals to technology-savvy users.

4 Data Portability Architectures

Several potential technical architectures exist that would help to economize on OSPs' transaction costs and be more user-friendly than the currently prevailing mode of indirect data portability. In the following, we present several technical architectures that enable direct data portability and discuss their implications from the perspectives of users, OSPs, and digital markets (Table 2).

4.1 OAuth Protocol and API

The OAuth protocol for authentication in combination with Application Programming Interfaces (APIs) for data transfer is frequently used to exchange data between two online services (Syrmoudis et al. 2021). Such combinations of APIs and OAuth authentication are offered by dominant OSPs such as Facebook, Google, and Apple or dedicated services such as Verimi. Often these solutions focus on data for login purposes, but it is typically up to these providers to define the extent to which they offer data exports. OSPs that decide to connect to one or more of these providers on their website can then use this connection to transfer basic personal data and optionally replace their own login method. As providers do not have to adhere to common standards, importing OSPs have to include each connection individually. While this approach allows users to transfer (limited) personal data more easily, providers offering authentication services may use it to gather additional data on users' behavior compromising their privacy and agency.

4.2 Data Portability Platforms

Dedicated platforms for direct data portability, such as the Data Transfer Project initiated by dominant OSPs like Google, Facebook, Microsoft, Twitter, and Apple, aim at facilitating bidirectional data transfer between participating

Table 2 Technical architectures enabling data portability

	Manual export and import	OAuth and API	Data portability platforms	Open protocols and service gateways	Self-hosting and personal data stores
Data portability approach	Indirect	Direct	Direct	Direct	Direct
Time	Up to 30 days	Real-time	Real-time	Real-time	Real-time
Frequency	Once (upon request)	Once (upon request)	Once (upon request)	Continuously[a]	Continuously
Scenario	Switch to another OSP, transfer data to complementary OSP	Transfer master data to complementary OSP	Transfer data to/between compatible online services	Exchange data with users of other online services, no switching to other OSP (multihoming)	Users own and control their data and selectively grant access to OSPs
Usability	**Low**: Manual user requests for export needed; user responsible for transfer to importing OSP	**High**: Only one click and login needed, but limited scope	**High**: Only one click and login needed	**High**: No change of OSP needed for connecting to other OSPs	**Low**: Complex setup and maintenance of personal data store
Scalability	**Very low**: Importing OSP needs to develop and maintain mechanisms for each supported OSP	**Low**: Importing OSP needs to adapt to API of each supported OSP	**High**: After connection to platform, data transfer from/to all connected OSPs possible	**High**: Data exchange is seamlessly possible between participating OSPs	**High**: Connection to data stores inherent part of ecosystem
Governance costs	**Low**: Minimal requirements for OSPs defined by legislator	**Low**: Exporting OSPs decide on individual implementation themselves	**Medium**: Central control and development of structure and data models	**High**: Development of common standards and protocols	**High**: Development of ecosystem, protocols, and common standards
Examples	User requests data from OSP and transfers received data to other OSP	Login with Facebook, Apple, Google, or Verimi	Data Transfer Project (DTP)	Federated Networks such as Mastodon using ActivityPub protocol, Matrix Bridges	Personal Online Data Stores such as Solid

[a]While the exchange of data between OSPs happens continuously, porting user data from one hosting provider to another still requires a regular data portability request

OSPs. It defines a set of "data models" that are standardized file formats and metadata. An OSP willing to participate has to develop two "adapters": An "authentication adapter", which could use OAuth, and a "data adapter" for transforming data to the format of the respective data model as defined by the Data Transfer Project. When users request a data transfer from OSP A to OSP B, they authenticate at both OSPs using the authentication adapters. The data from OSP A is then transformed to the data model using its data adapter and subsequently transformed to OSP B's data formats using OSP B's data adapter. This approach makes it easy for users to request a data transfer between participating OSPs. For the providers themselves, participating in a project like the Data Transfer Project can substantially lower the cost for implementing data portability as developing the two adapters suffices to connect to all other participating OSPs. However, without regulatory oversights, these consortia may be dominated by OSPs with greater market power and resources, which may opportunistically exploit their powerful position to specify standards serving their strategic and technical purposes.

4.3 Open Protocols and Service Gateways

Apart from methods that require users to request a one-time transfer of data, there are also solutions for a continuous transfer of data. The usage of open protocols and service gateways is a way to enable interoperability between services operated by different OSPs. With interoperability, users do not have to switch to another OSP with which they want to interact but can do so using their existing account. Solutions for interoperability either require two or more OSPs to use a common (open) protocol or one OSP to develop a gateway which parses data from other OSPs in real time. A prominent example for such a gateway is a Matrix bridge. For instance, they can be used to connect (group) chats which utilize the Matrix protocol to other OSPs (e.g., Slack) or other protocols and can transfer data and messages between OSPs in real time. A bridge acts as a hidden intermediary that reads data from one OSP and sends it to another OSP in real-time and vice-versa. Connected OSPs do not need to cooperate or know of the existence of a bridge. While bridges are a way to implement data transfers and compatibility between OSPs, bridging can violate OSPs' terms and conditions and negatively affect the privacy of users when they are not informed that their data is processed via a bridge.

Another option for implementing interoperability for two or more OSPs that want to exchange data in real time are common protocols. By using a common protocol like the "ActivityPub" protocol (Lemmer-Webber et al. 2018), OSPs can "federate" and allow their users to interact with each other. OSPs have to allow federated OSPs to access their data from a standardized API and vice versa read data of federated OSPs from their APIs. The overall network architecture defines how data is transferred. In case of ActivityPub, users and servers have standardized inboxes and outboxes. Messages and other data can be read from the own inbox, sent to other inboxes (allowing only the specified user/server to read it), sent to the own outbox (allowing everyone to read it), and read from other outboxes. However, depending on the actual implementation, interoperability can negatively impact OSPs' ability to innovate. OSPs might need to adhere to unfavorable standards and protocols to comply and protocol changes need to be implemented by all parties or be downwards compatible, which may slow down the rate of innovation.

4.4 Self-Hosting and Personal Data Stores

Another alternative for transferring data between OSPs is the separate hosting or self-hosting of data controlled by users known as "personal online data stores" (Capadisli et al. 2021; Mager and Kranz 2021). Instead of OSPs storing user data on servers they control, this architectural approach puts users in control of their data which is hosted by an entity other than the OSP in question (i.e., users themselves or third-party providers). Thus, service provision and data ownership would be separated which would increase competition based on service quality and lower the importance of advantages gained by owning vast amounts of data (Sunyaev et al. 2021). Beyond that, initiatives such as the Swiss Data Alliance[6] provide guidance for implementing data portability as they extend the storage of personal online data to an open and shared data repository that includes data from government, businesses, research, education and culture.

5 Discussion and Implications

With the right to data portability, regulators aim at improving users' privacy, choice, and options to control and reuse their data, leading to more transparency, data-based innovation and competition, and eventually "data-richness" in online service markets (Jarke et al. 2019). Greater control and fluidity of data reduce users' switching costs and therefore lock-in effects that cement the dominance of a few OSPs. Thus, RtDP should increase competition and the rate of innovation as unlocking data from proprietary silos and increasing users' rights to control data counters the unfavorable skewed allocation of data that currently limits opportunities for data-based innovation (Gregory et al. 2021; Jones and Tonetti 2020).

[6] See: https://www.swissdataalliance.ch/

Despite RtDP's inherent promises, RtDP has not yet had a major influence on digital markets. We have outlined that the main reasons are on the levels of users, markets, and technical architectures. Hence, we need research that addresses these barriers and uncertainties to better understand the (un-)intended consequences of different RtDP implementations on stakeholders. In the following, we outline key avenues for future research in BISE/IS and adjacent disciplines and summarize them in Table 3.

5.1 User Level

Given the perils of ''surveillance capitalism'' and the opportunities of ''data openness'', we need to improve our understanding on how to motivate users to play a more active role in data markets and what level of data sharing is most beneficial for users (Alt et al. 2021; Zuboff 2019). In this regard, we propose two important avenues for future research on the level of users.

What motivates users to actively self-manage their data and make use of the RtDP?

Although many internet users state that they are generally concerned about their online privacy, many fail to act accordingly and to effectively self-manage their privacy settings (Acquisti et al. 2020; Adjerid et al. 2018). As such, usage of the RtDP is low. Even when users plan to switch to another OSP, they do not consider making use of the RtDP, mostly due to lack of awareness and concerns about loss of information (Luzsa et al. 2022b). Several studies have examined this behavioral privacy contradiction summarized as the privacy paradox (e.g., Acquisti et al. 2020; Adjerid et al. 2018), as well as the impact of nudging on users' privacy behaviors (Mager and Kranz 2021) or users' intentions based on their threat and efficacy perceptions (e.g., Johnston et al. 2015). We, therefore, need to improve and extend our knowledge on how to counteract the privacy paradox and motivate users to actively self-manage their data and privacy settings. Thus, we call for design science research and field experiments that address the design, implementation and evaluation of self-management privacy settings and requests, closely involving and integrating users' perspectives and requirements with a particular focus on data portability. Further research should explore how dark patterns can be effectively avoided (Acquisti et al. 2017) and how users' intention to self-manage their data is influenced by different motivational processes. Moreover, research should investigate the extent to which existing privacy regulations that permanently require users to engage with their privacy settings contribute to privacy fatigue – a state of emotional exhaustion and cynicism (Choi et al. 2018) – and how to establish comprehensive privacy regulations that ease users' burden of the self-management of their data (Acquisti et al. 2020).

This line of work should also explore the development and usage of tools for users to actively explore and manage exported data. On the one hand, understanding their data and carefully selecting data for data imports enhances users' general awareness of data practices in the digital economy and offers obvious privacy benefits. On the other

Table 3 Summary of future research directions

Level of analysis	Research questions	Suitable BISE/IS research streams	Foundational literature
User	What motivates users to actively self-manage their data and make use of the RtDP?	Digital nudging, online privacy, user motivation, privacy fatigue	Acquisti et al. (2020); Choi et al. (2018); Johnston et al. (2015); Mager and Kranz (2021)
	What are the effects of different data scopes and architectures on user adoption and usage of data portability?	Data governance, exchanges, markets; IS adoption and continuance	De Hert et al. (2018); Krämer (2020); Wohlfarth (2019)
Markets	Which portability implementations are most beneficial for stakeholders and society and what is the role of boundary conditions?	Multihoming, network effects, digital platforms, gatekeepers; individual data ownership	Easley et al. (2018); Lam and Liu (2020); Ramos and Blind (2020); Sunyaev et al. (2021)
	How efficient and promising is the approach of 'in situ' data rights?	IS economics, data exchanges, federated learning mechanisms	Agrawal et al. (2021); Bonawitz et al. (2019); Van Alstyne et al. (2021)
Technical	How can technical requirements be refined and what common standards need to be defined to make the RtDP an effective user right?	Technology standard making, network effects, multihoming	Willard et al. (2018); Wong and Henderson (2019)
	How can data portability solutions be developed to converge towards ecosystems with interoperable OSPs?	Ecosystem governance, federated networks	Capadisli et al. (2021); Lemmer-Webber et al. (2018)

hand, data editing may also influence the meaningfulness and veracity of data.

What are the effects of different data scopes and architectures on user adoption and usage of data portability?

Current data portability regulations stipulated in the GDPR and Data Act restrict the scope of personal data to *received* and (broadly interpreted) *observed data* (De Hert et al. 2018; Krämer 2020). This restriction limits the right's effectiveness, as *inferred* and *predicted* personal data derived from data provided by the users is currently excluded, while these data types are particularly relevant for innovative business models and harbor significant privacy implications. Hence, we need studies that investigate how different data scopes relate to user adoption, service quality, competition, and innovation. Research should also investigate and design solutions for the transfer of sensitive personal data, such as social security numbers, financial information, or health data, which bears significant privacy and security risks (Krämer 2020; Wohlfarth 2019).

Likewise, the current scope of the regulation limits the applicability of the RtDP in the context of third parties. Entities such as data brokers and stakeholders in the technical advertising ecosystem extensively draw on users' personal data, but may not have been "provided" with this data directly by the user. The question on how to communicate the current limits of regulation to users, while generally raising awareness of the benefits of data portability, therefore, constitutes another key challenge.

5.2 Market Level

The biggest beneficiaries of improved availability of user data through data portability should be rivals of dominant OSPs and new entrants. In an effective data portability regime, they could 'absorb' user data from dominant OSPs, which would improve their capabilities to innovate and overcome competitive barriers such as lock-in or data network effects (Wohlfarth 2019). As a result, dominant OSPs have increased incentives to invest in data-driven innovation and improve existing and currently developing new technologies in order to sustain their competitiveness and prevent user churn (Lam and Liu 2020; Ramos and Blind 2020). However, it is difficult to determine the actual economic impact of the RtDP given its low adoption and limited experiences with data portability. So far, only the example of mobile number portability exists that showed that achieving the desired impact in a market with vested interests and opposing incentives is complex – even though only highly standardized data needs to be transferred among a limited number of market players (Maicas et al. 2009). Consequently, further research should address the following questions.

Which portability implementations are most beneficial for stakeholders and society and what is the role of boundary conditions?

The actual impact of data portability on innovation and market competition will depend on several boundary conditions, most importantly, data scope and quality, duration, recency and frequency of data transfers, inherent value of different data types, and the strength of specific markets' network effects (Krämer 2020; Lam and Liu 2020; Ramos and Blind 2020). For instance, considering that network effects are stronger for social networks than search engines, the RtDP will likely have a stronger effect on social network services than search engines. Likewise, we assume that users of social network services in comparison to search engines are more likely to multihome. Multihoming reduces the market concentration of the few dominant OSPs as users join multiple providers simultaneously (Ramos and Blind 2020). However, the impact of the RtDP on user switching (i.e., users transfer data to another service and terminate previous service usage) vis-à-vis its impact on multihoming (i.e., users transfer data to another service, but keep using the previous service), needs further investigation. Moreover, the Digital Markets Act's increased portability obligations for gatekeepers need to be evaluated regarding their potential effects on market competition, innovation, and welfare.

The current approaches to data portability in terms of transferring personal data from one OSP to another may be associated with several unforeseen disadvantages. Transferred data may lose its context and algorithms can no longer access, compare, and analyze other users' personal data of the original OSP (Van Alstyne et al. 2021). Data will no longer stay current as it will not be constantly updated and transferred data will have to be reconnected and'reanimated' first to be acted upon. Furthermore, data exports enable data editing and hence data falsification, which may lead to market failures due to moral hazard, reduced data network effects, and slowed innovation (Gregory et al. 2021; Van Alstyne et al. 2021). As a result, future research is needed on whether the current approach to data portability is best suited to promote competition and ensure users' control over their online privacy.

Several alternative approaches have, therefore, been put forth. The concept of *separate hosting of data* builds on the notion that OSPs do no longer control user data, but users themselves have control to manage their data generated through OSP usage. Consequently, data storage is disentangled from data-based services and OSPs need explicit user permission to be able to access their data (Jones and Tonetti 2020; Sunyaev et al. 2021). However, the effectiveness and multi-level effects of these implementation approaches need to be better understood.

How efficient and promising is the approach of 'in situ' data rights?

The aim of 'in situ' data portability is to keep data in its location to avoid unintended consequences of 'ex situ' data portability, and to "*bring the algorithms to the data [*in situ*] instead of bringing the data to the algorithms [*ex situ*]*" (Van Alstyne et al. 2021). 'In situ' data rights may have several benefits in comparison to 'ex situ' data portability, such as data keeping its contextual value, remaining up to date, and reducing the risk of data falsification (Van Alstyne et al. 2021). Thus, future research needs to analyze the extent to which this approach can work hand-in-hand with technical privacy-preserving measures (e.g., Agrawal et al. 2021). Further, we suggest investigating potential implementations of 'in situ' rights, such as through federated learning mechanisms, which enable model training on decentralized data through distributed machine learning (Bonawitz et al. 2019). For instance, gathering and curating data from several different sources at shared platforms – *data exchanges or data spaces* – enables algorithms to be trained locally ('in situ') in these shared data repositories. Moreover, individuals and organizations contributing to a data exchange or data space can further benefit from the value of the aggregated data, since a collective data exchange platform can sell these data as information at an adequate price. However, these data exchanges may become a novel equivalent of data monopolies. In comparison to traditional platform gatekeepers that connect OSPs with users and exercise control of the data services running over their platform through data neutrality (Easley et al. 2018), data exchanges exercise control over the algorithms running on their data. Hence, the consortium-driven approach of open data spaces such as GAIA-X may prove more effective to prevent concentration of market power and lock-in effects (Otto and Jarke 2019).

5.3 Technical Implementation

In practice, data portability is only possible to a very limited extent and specific technical standards in relation to data formats' conformity and implementation are missing. For indirect data portability, there are no precise specifications on how OSPs have to export data and, more importantly, providing documentation is not mandatory. Regarding direct data portability, the perfect solution that is suitable for all use cases does not exist. Industry consortia such as the Data Transfer Project could ease the transfer of personal data between OSPs and therefore allow users to switch their OSPs more easily. However, as long as regulators do not mandate or foster the development of direct data portability platforms, these platforms may be dominated by large OSPs who can set the speed of development and build the architecture in a way that is most favorable for them. Therefore, we suggest further research to address the following key questions.

How can technical requirements be refined and what common standards need to be defined to make the RtDP an effective user right?

Research has shown that the technical requirements mandated by the RtDP are too unspecific and can be fulfilled without adhering to common standards (Wong and Henderson 2019). Furthermore, OSPs do not have to provide documentation on their data export practices, which would facilitate data import for other OSPs. These technical shortcomings may also contribute to the reluctance of OSPs to offer import options (Syrmoudis et al. 2021). Amending the RtDP by a provision which obliges OSPs to provide public documentation on the structure of their exports or to standardize them would facilitate data imports. To enable user-friendly, secure, and effective data portability between OSPs, research is needed that analyzes how effective standards could be developed and implemented to avoid lock-ins to inferior standards (Willard et al. 2018; Wong and Henderson 2019). Further, we need to better understand which standardization processes (e.g., de jure, de facto) and approaches (management-based, technology-based, or performance-based standards) are most effective and efficient and how different standard options will impact competition and innovation in digital service markets and between stakeholders (Zeiss et al. 2021). Factors to consider include network effects, multihoming, standard adoption, standardization costs, and social welfare.

How can data portability solutions be developed to converge towards ecosystems with interoperable OSPs?

Especially in scenarios with network effects, concepts where data does not have to be hosted by an OSP to connect to its services can be a feasible solution. When OSPs are interoperable, users have more freedom to choose their hosting provider and do not need to have their personal data stored by multiple OSPs. However, making services interoperable or designing interoperable ecosystems induces a high effort in developing standards and protocols as well as posing the additional risk of limiting their adaptability after implementation. It is an open question of how to design service ecosystems that are interoperable while allowing participating OSPs to remain innovative. Feasible approaches with similar goals that are under development include federated networks (Lemmer-Webber et al. 2018) and service ecosystems where data is stored separately from the provider of an online service (Capadisli et al. 2021).

In a similar vein, the Digital Markets Act aims at interoperability and continuous data transfer. The effects of these new mandates for gatekeepers to provide continuous,

real-time access to data and enable interoperability with their operating system, hardware, or software features will need to be closely monitored and investigated. In comparison to the RtDP, the new mandates move closer to enabling interoperability, although the regulations only apply to gatekeepers and do not explicitly envision a reciprocal exchange between gatekeepers and other OSPs. However, to truly empower users in juxtaposition to OSPs and gatekeepers, users need to be able to transfer their data continuously and in real-time and to a diverse set of OSPs (Krämer 2020).

6 Conclusion

Our study aimed at increasing the conceptual clarity of the data portability concept and providing an analysis of current and potential implementations and their effects. We further discuss the inherent potential, promises, and challenges of data portability in relation to the BISE/IS community and highlight avenues for future research. While many questions remain on how to enable and promote the effective use of data portability, we believe that the concept has the potential to address apparent market failures in digital markets by facilitating more competition and data-driven innovation. To make data portability a meaningful user right and an effective factor for the contestability of digital markets, continuous research is needed that analyzes the (unintended) consequences and helps fine-tune data portability regulations and practices. Our article intends to contribute to the discussion and to make data portability an effective user right.

Acknowledgements We would like to thank the anonymous reviewers for helpful feedback. We further are grateful for funding support from the Bavarian Research Institute for Digital Transformation (bidt). Responsibility for the content of this publication rests with the authors.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Acquisti A, Adjerid I, Balebako R, Brandimarte L, Cranor LF, Komanduri S, Leon PG, Sadeh N, Schaub F, Sleeper M (2017) Nudges for privacy and security: understanding and assisting users choices online. ACM Comput Surv 50(3):1–41

Acquisti A, Brandimarte L, Loewenstein G (2020) Secrets and likes: the drive for privacy and the difficulty of achieving it in the digital age. J Consum Psychol 30(4):736–758

Adjerid I, Peer E, Acquisti A (2018) Beyond the privacy paradox: objective versus relative risk in privacy decision making. MIS Q 42(2):465–488

Agrawal N, Binns R, Kleek MV, Laine K, Shadbolt N (2021) Exploring design and governance challenges in the development of privacy-preserving computation. In: Proceedings of the 2021 CHI conference on human factors in computing systems, Yokohama, Article 68. https://doi.org/10.1145/3411764.3445677

Alt R, Göldi A, Österle H, Portmann E, Spiekermann S (2021) Life engineering. Bus Inf Syst Eng 63(2):191–205

Autor D, Dorn D, Katz LF, Patterson C, Van Reenen J (2020) The fall of the labor share and the rise of superstar firms. Q J Econ 135(2):645–709

Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, Kiddon C, Konečný J, Mazzocchi S, Mcmahan B (2019) Towards federated learning at scale: system design. In: Proceedings of the 2nd SysML conference, 1, pp 374–388

Capadisli S, Berners-Lee T, Verborgh R, Kjernsmo K (2021) Solid protocol. https://solidproject.org/TR/protocol. Accessed 29 Apr 2023

Choi H, Park J, Jung Y (2018) The role of privacy fatigue in online privacy behavior. Comput Hum Behav 81:42–51

De Hert P, Papakonstantinou V, Malgieri G, Beslay L, Sanchez I (2018) The right to data portability in the gdpr: towards user-centric interoperability of digital services. Comput Law Secur Rev 34(2):193–203

Easley RF, Guo H, Kraemer J (2018) Research commentary – From net neutrality to data neutrality: a techno-economic framework and research agenda. Inf Syst Res 29(2):253–272

Engels B (2016) Data portability among online platforms. Internet Policy Rev 5(2):1–17

European Data Protection Board (2022) Guidelines 01/2022 on data subject rights–right of access. https://edpb.europa.eu/system/files/2022-01/edpb_guidelines_012022_right-of-access_0.pdf. Accessed 29 Apr 2023

Fadler M, Legner C (2022) Data ownership revisited: clarifying data accountabilities in times of big data and analytics. J Bus Anal 5(1):123–139

Gregory RW, Henfridsson O, Kaganer E, Kyriakou H (2021) The role of artificial intelligence and data network effects for creating user value. Acad Manag Rev 46(3):534–551

Jarke M, Otto B, Ram S (2019) Data sovereignty and data space ecosystems. Bus Inf Syst Eng 61(5):549–550

Johnston A, Warkentin M, Siponen M (2015) An enhanced fear appeal rhetorical framework: leveraging threats to the human asset through sanctioning rhetoric. MIS Q 39(1):113–134

Jones CI, Tonetti C (2020) Nonrivalry and the economics of data. Am Econ Rev 110(9):2819–2858

Krämer J (2020) Personal data portability in the platform economy: economic implications and policy recommendations. J Compet Law Econ 17(2):263–308

Kuebler-Wachendorff S, Luzsa R, Kranz J, Mager S, Syrmoudis E, Mayr S, Grossklags J (2021) The right to data portability: conception, status quo, and future directions. Informatik Spektrum 44:264–272

Lam WMW, Liu X (2020) Does data portability facilitate entry? Int J Ind Organ 69:102564

Lemmer-Webber C, Tallon J, Shepherd E, Guy A, Prodromou E (2018) Activitypub [W3C Recommendation]. https://www.w3.org/TR/activitypub/. Accessed 29 Apr 2023

Luzsa R, Mayr S, Syrmoudis E, Grossklags J, Kübler-Wachendorff S, Kranz J (2022a) Online service switching intentions and attitudes towards data portability – the role of technology-related attitudes and privacy. In: Mensch und computer 2022a, Darmstadt. https://doi.org/10.1145/3543758.3543762

Luzsa R, Mayr S, Syrmoudis E, Grossklags J, Kuebler-Wachendorff S, Kranz J (2022b) Datenportabilität Zwischen Online-Diensten. Nutzeranforderungen und Gestaltungsempfehlungen. Ergebnisse einer Bevölkerungsrepräsentativen Befragung. [Data portability between online services. user requirements and design recommendations. Results of a population-representative survey]. bidt, Working Paper 5. https://doi.org/10.35067/bv16-2z31

Mager S, Kranz J (2021) Consent notices and the willingness-to-sell observational data: evidence from user reactions in the field. ECIS 2021 Research Papers. 89. https://aisel.aisnet.org/ecis2021_rp/89. Accessed 29 Apr 2023

Maicas JP, Polo Y, Sese FJ (2009) Reducing the level of switching costs in mobile communications: the case of mobile number portability. Telecommun Policy 33(9):544–554

Otto B, Jarke M (2019) Designing a multi-sided data platform: findings from the international data spaces case. Electron Mark 29(4):561–580

Pavlou PA (2011) State of the information privacy literature: where are we now and where should we go? MIS Q 35(4):977–988

Ramos EF, Blind K (2020) Data portability effects on data-driven innovation of online platforms: analyzing spotify. Telecommun Policy 44(9):102026

Rupp E, Syrmoudis E, Grossklags J (2022) Leave no data behind – empirical insights into data erasure from online services. In: Proceedings on Privacy Enhancing Technologies 2022(3):437–455

Sunyaev A, Kannengießer N, Beck R, Treiblmaier H, Lacity M, Kranz J, Fridgen G, Spankowski U, Luckow A (2021) Token economy. Bus Inf Syst Eng 63(4):457–478

Syrmoudis E, Mager S, Kuebler-Wachendorff S, Pizzinini P, Grossklags J, Kranz J (2021) Data portability between online services: an empirical analysis on the effectiveness of GDPR Art 20. Proc Priv Enhanc Technol 3:351–372

Van Alstyne MW, Petropoulos G, Parker G, Martens B (2021) "In situ" data rights. Commun ACM 64(12):34–35

Willard B, Chavez J, Fair G, Levine K, Lange A, Dickerson J (2018) Data transfer project: from theory to practice. https://services.google.com/fh/files/blogs/data-transfer-project-google-whitepaper-v4.pdf. Accessed 29 Apr 2023

Wohlfarth M (2019) Data portability on the internet. Bus Inf Syst Eng 61(5):551–574

Wong J, Henderson T (2019) The right to data portability in practice: exploring the implications of the technologically neutral GDPR. Int Data Priv Law 9(3):173–191

Zeiss R, Ixmeier A, Recker J, Kranz J (2021) Mobilising information systems scholarship for a circular economy: review, synthesis, and directions for future research. Inf Syst J 31(1):148–183

Zuboff S (2019) The age of surveillance capitalism: the fight for a human future at the new frontier of power. Profile, London