sciendo

Proceedings on Privacy Enhancing Technologies ; 2021 (3):351–372

Emmanuel Syrmoudis*, Stefan Mager, Sophie Kuebler-Wachendorff, Paul Pizzinini, Jens Grossklags, and Johann Kranz

Data Portability between Online Services: An Empirical Analysis on the Effectiveness of GDPR Art. 20

Abstract: Data portability regulation has promised that individuals will be easily able to transfer their personal data between online service providers. Yet, after more than two years of an active privacy regulation regime in the European Union, this promise is far from being fulfilled. Given the lack of a functioning infrastructure for direct data portability between multiple providers, we investigate in our study how easily an individual could currently make use of an indirect data transfer between providers. We define such porting as a two-step transfer: firstly, requesting a data export from one provider, followed secondly by the import of the obtained data to another provider. To answer this question, we examine the data export practices of 182 online services, including the top one hundred visited websites in Germany according to the Alexa ranking, as well as their data import capabilities. Our main results show that high-ranking services, which primarily represent incumbents of key online markets, provide significantly larger data export scope and increased import possibilities than their lower-ranking competitors. Moreover, they establish more thorough authentication of individuals before export. These first empirical results challenge the theoretical literature on data portability, according to which, it would be expected that incumbents only complied with the minimal possible export scope in order to not lose exclusive consumer data to market competitors free-of-charge. We attribute the practices of incumbents observed in our study to the absence of an infrastructure realizing direct data portability.

Keywords: Data portability, Privacy regulation, Competition between online services, General Data Protection Regulation (GDPR), Data economy, Consumer rights, Switching costs, Data controller

DOI 10.2478/popets-2021-0051
Received 2020-11-30; revised 2021-03-15; accepted 2021-03-16.

***Corresponding Author: Emmanuel Syrmoudis:** Technical University of Munich, E-mail: emmanuel.syrmoudis@tum.de

1 Introduction

Personal data is a nonrival good and thus in theory can be consumed indefinitely [1]. Yet, in contrast to consumption, the collection of data in many key online services such as social networks, map or fitness applications is performed in a rival manner [2]. Individuals will preferably only use one map or fitness application for informational support in their everyday life.

As data has emerged to be a critical competitive resource in the early 21st century [3, 4], this rivalry in data collection provides strong incentives for dominant platforms that offer key online services to lock in consumer data and to deny individuals and competitors alike access to this data as a knowledge-building resource [1]. This corporate behavior of data siloing results in two main societal challenges [5, 6]: one of concentrated epistemological power, i.e., who has the right to know the most and control accumulated information about a person [7], and another challenge of concentrated economic power, i.e. who can process the accumulated information about a large number of individuals in order to best know their needs and innovate [1].

Data portability regulation inherently promises to function as one possible remedy against the concentration of both kinds of power in too few hands [8]. If applied in an appropriate way, it will enable individuals to gain greater informational self-determination on the one hand, and act as a continuous impulse for competition in digital markets on the other [9]. However, while

Stefan Mager: Ludwig-Maximilians-University of Munich, E-mail: stefan.mager@lmu.de
Sophie Kuebler-Wachendorff: Ludwig-Maximilians-University of Munich, E-mail: kuebler-wachendorff@bwl.lmu.de
Paul Pizzinini: Ludwig-Maximilians-University of Munich, E-mail: pizzininipaul@gmail.com
Jens Grossklags: Technical University of Munich, E-mail: jens.grossklags@in.tum.de
Johann Kranz: Ludwig-Maximilians-University of Munich, E-mail: kranz@lmu.de

previous studies have empirically analyzed compliance with the *General Data Protection Regulation (GDPR)* [10], or have even started to investigate the data export execution of the GDPR's Right to Data Portability (Art. 20) [11], thus far, we have been unable to locate any empirical study analyzing actual data transfer between providers. We thus pose the research question: **How well does data portability regulation currently deliver on its promise to allow individuals to transfer data from one provider to another?**

This main research question of our study will be examined through a series of subquestions: firstly, in the absence of a functioning and comprehensive infrastructure for direct data transfer between data controlling services[1] [12], we examine the question of whether online services provide individuals with an option to easily export data and furthermore with the possibility to import data. If yes, which options do they offer? Secondly, we investigate what scope of personal data online services provide via data exports by building on the data taxonomy of De Hert et al. [9]. Drawing on Wohlfarth [4], we would expect that current market-leading services would have an incentive to provide data subjects only with the minimum scope possible in order to maintain their advantage in the rivalry of data collection over new entrants or smaller competitors. Thirdly, we examine how quickly, securely, and to what level of compliancy with GDPR Art. 20, personal data is transmitted to individuals upon request. In this way, we can observe whether there has been any improvement in comparison to prior work of Wong and Henderson [11] regarding the ease of data transfer.

We examine the personal data export and import capabilities of 182 online services, among which are the 100 most visited websites in Germany according to the Alexa ranking of September 2020. Using the Alexa rank as proxy for the popularity of a service, we conduct a regression analysis to investigate whether it plays a decisive role in how well data controllers implement their legal requirement to provide data portability to individuals. Based on our sample, our study further investigates whether certain industry sectors enable a more effective portability of data between providers when compared with other industry sectors. Our categorization relies on the NACE Rev. 2 industry classification, which is the official industry classification of the European Union [13].

Within this empirical study, we find evidence that popular services with higher Alexa ranks complied significantly more often with GDPR Art. 20, when responding to a data portability request. Contrary to what we expected from the theoretical literature on data portability and the economics of data, popular services with higher Alexa ranks also provided a higher data scope than their competitors with lower ranks. Lastly, competitors to those market leading services offer consumers significantly fewer data import possibilities and thus have not yet started to leverage the opportunity of lowered switching costs through data portability regulation for themselves, which potentially could help them to win consumers by means of indirect data portability.

Our study has several implications for policy makers. Although the introduction of data portability regulation was designed to be a step towards maintaining a level playing field in online markets, we cannot find significant evidence for these intended effects regarding direct (GDPR Art. 20 (2)) or indirect data portability (Art. 20 (1)). In order for privacy regulations to generate more economic implications for online services, regulators should make entrepreneurs and small service providers more aware of the possibilities privacy regulations offer them in winning new users. To achieve this, projects that enable the direct transfer of personal data between providers, such as the DTP [14], should be promoted. Additionally, penalties for inadequate handling of data portability legislation should be enforced so that corporations are pressurized to act and participate in projects that aim to establish the direct and secure transfer of personal data between providers.

The remainder of this paper is structured as follows: in Section 2, we provide a review of both theoretical work and empirical investigations in the academic literature, as well as recent legislative developments on data portability. In Section 3, we describe in detail how we collected our dataset and analyzed the export and import capabilities of the most important online services in Germany. In Section 4, we display the summary statistics and describe the results of our regression analyses, before we proceed to discuss their implications for research and policy in Section 5. We close our study by mapping out avenues for future research on data portability in Section 6 and end with a brief summary of the major contributions and limitations of our research in the conclusion section.

1 The Data Transfer Project (DTP) is perhaps the most advanced project to provide a direct data portability infrastructure between online services. However, as of 2020 it is still in beta mode and has a low number of participating services.

2 Background

2.1 Data portability in privacy legislation and in practice

In May 2018, the European Union introduced a Right to Data Portability (RtDP) for personal data enshrined for the first time in a legislative framework [15]. The right consists of two subrights, which allow European citizens to either receive and store personal data concerning themselves on a storage medium of their choice (Art. 20 (1) GDPR) or to transfer their personal data directly to another provider on request (Art. 20 (2) GDPR). In particular due to the second subright, the RtDP has the potential to serve as a remedy against user lock-in to online services - or data controllers, as they are referred to in legislative texts - by lowering the switching costs of consumers [8, 16].

In the subsequent years, a number of jurisdictions, as for example California with its *California Consumer Privacy Act* (CCPA), Brazil with its *Lei Geral de Proteção de Dados Pessoais* (LGPD) or India with its *Personal Data Protection Bill* (PDPB), have followed the lead of the European Union and have adopted or are about to adopt a comprehensive privacy regulation[2]. These three privacy regulations partly determine a RtDP similar to the GDPR with both subrights (LGPD, PDPB) or partly just integrate the first subright of the GDPR's RtDP in the respective Right of Access (RoA) of their framework (CCPA) [17].

In general, a Right of Access (RoA), Right to Rectification (RtR), and Right to Erasure (RtE) of personal data are central user privacy rights in all of the newly established privacy regulations aside from the RtDP. Each of these rights aims to raise the bargaining power of consumers in digital markets [18] and to increase consumers' informational self-determination in comparison to pre-GDPR privacy laws. Regrettably, prior work [19–22] demonstrated – on the basis of a previous Californian privacy law – that corporations frequently ignored or declined requests of individuals for information access or corporate information sharing practices. As such, the RtDP – and especially its subright of direct data transfer between providers – represents the right with the highest potential economic implications *only if* it is applied correctly [9].

However, determining (and verifying) the proper implementation of the RtDP is a task that legislators have left to consumers, data protection organizations, and data controllers. After a preliminary tug-of-war between EU policy makers and lobbyists of online services during the enactment of the GDPR [23], its implementation is envisioned to evolve as a result of a common collaborative process [9].

2.2 Models on the economics of data portability

Existing academic literature evaluates data portability implementation based on theoretical mathematical modelling. These analyses argue that the amount of personal data collected by online services determines their market competitiveness as well as the size of the user lock-in effect [4]. This effect lowers the incentive for users to move from one service provider to another, as the costs for setting up a new profile with a comparable amount of data will be higher. Examples for these switching costs can be observed in email or digital storage services, where besides general individual profile information, other documents such as files, photos or contacts need to be re-uploaded, or online banking accounts because bank transfers need to be re-entered [4]. Therefore, the RtDP is designed to counteract user switching costs and to increase competition among online services [24].

Economic incentives for online service providers to comply with the RtDP may well be limited. Having control over a vast amount of data about individuals and their habits provides a variety of possibilities for monetization, commodification, and control [7]. Comparing users' needs to online services' economic incentives highlights the delicate balance between protecting consumer rights on the one hand, and preserving the intellectual property rights of services on the other hand [25].

According to Wohlfarth [4], further issues might arise when porting data, such as privacy and security concerns from transfering sensitive data (e.g., credit card details or social security numbers). Additionally, in a regulatory regime with a RtDP, a general increase in collected data is to be expected according to Wohlfarth [4], as market entrants now need a greater amount of personal data to be able to provide a competitive service quality. New market entering services are expected to be the general beneficiaries of a data portability reg-

2 The CCPA and LGPD have taken effect as of January 2020 and August 2020, respectively. As of December 2019, the PDPB is in draft form.

ulation due to improved access to user data from incumbent services, when users port their data. However, the increased incentives to collect an even greater amount of data come at the users' expense [4].

Lam and Liu [26] argue that the promise of data portability may actually limit its impact in practice. Believing in the future option to transfer data to a competitor, consumers may even contribute more data, and thereby strengthen the competitive advantage of the incumbent, in particular, when advanced data analytics provide a better and more customized service.

Krämer and Stüdlein [24], in contrast to both [4] and [26], focus on two strategic variables instead of just one in their market evaluation: price and disclosure level. Their results show that the RtDP will overcome the incentive of the incumbent service to disclose little user data, since data can now be ported free-of-charge. This strengthens the competitiveness of entrants and thus leads to a reduction in service price that the incumbents can charge, whereas the service price of the entrants increases [24]. Consumers of incumbent online services as well as consumers switching from an incumbent to an entrant are therefore ultimately better off. In contrast, consumers who were already with the entrants before the RtDP are strictly speaking worse off, due to lower quality compared to incumbents as well as higher prices compared to the market situation without the RtDP [24, 26].

2.3 Requirements on scope and data format

Art. 20 (1) states that a "data subject has the right to receive the personal data concerning him or her, which he or she has provided to a controller" [27]. De Hert et al. [9] analyze from a legal perspective what scope of data online services have to provide upon a RtDP request. They distinguish four possible types of data scope: data directly submitted by the user (*received data*), *observed data* which includes data gathered by sensors, e.g., location data, and *inferred data* and *predicted data*, which both are created by data controllers on the basis of received and observed data.

De Hert et al. [9] develop a restricted interpretation of the RtDP, favored by the semantic meaning of the term *provided* in Art. 20 (1), where only received data needs to be made available to the user. An alternative, extensive interpretation which aligns with the aim of the RtDP to give users more control over their data, includes both received and observed data. Inferred and predicted data never have to be provided in response to a RtDP request according to their analysis, since they already represent intellectual work of the data controlling online service provider [28].

Regarding the data format, Art. 20 (1) states that the data has to be provided "in a structured, commonly used and machine-readable format". This restricts the services in the data formats they are allowed to use and aims to give users the data in a format which easily enables them to make further use of their data.

In the *Guidelines on the Right to Data Portability* [29], it is clarified that the desired outcomes of the right to data portability are "interoperable systems, not compatible systems". Thus, the used data format should be suitable for achieving the goal of interoperability. The data format is expected to have a "high level of abstraction from any internal or proprietary format". More specifically, the guidelines state that the data should be provided in a "commonly used open format [...] along with useful metadata at the best level of granularity" [29]. They name XML, JSON, and CSV as possible formats that fulfill the criteria.

It remains an open question as to how exactly "structured" and "machine-readable" should be interpreted. Furthermore, while formats like XML and JSON have an inherent structure, a file where all data is stored as continuous text using only one key-value pair is still a valid XML/JSON file, but its content cannot be accessed in a structured way.

Table 1. Common data formats and their characteristics [11].

	Structured	Commonly used	Machine-readable
CSV	yes	yes	yes
HTML	?	yes	?
JPEG	no	yes	no
JSON	yes	yes	yes
PDF	?	yes	?
XLS	yes	yes	?
XML	yes	yes	yes

Wong and Henderson [11] provide an extensive overview on the characteristics of different file formats. They assess which formats have the potential to be compliant with Art. 20 using recommendations of the Information Commissioner's Office of the UK. Table 1 lists their assessments for seven common file formats. Entries marked with "?" are assessed as ambiguous.

3 Methodology

3.1 Data collection

Contrary to what GDPR Art. 20 (2) envisions, no functioning infrastructure for direct transfer of data between online services exists as of today. Given this status quo, the economic impact of the GDPR's Right to Data Portability is naturally limited. Yet, an indirect transfer in which a person first exports their personal data from one provider and then offers it for import to another provider should be a feasible option to an individual who wants to change their provider of a certain online service. Thus, we examine current data export and import capabilities of online services which ranked highest on the Alexa website ranking to assess the present state of data portability in practice.

Three authors of this article, who live in Europe, made data portability requests to 182 online services with reference to GDPR Art. 20 (1) between January and September 2020. Similar to Wong and Henderson [11], they contacted the online services via email (see Appendix A) in case no automated download option or portability contact request form was available on the service's website. They then monitored various process characteristics of the providers' responses: the success of the request, the days to fulfillment, the number of actions needed to obtain one's personal data, the authentication requirements, the transmission path chosen to send the personal data, confusions with other GDPR rights, as well as the data types and file formats provided. (We will discuss these statistics in detail in Section 4.2.)

The online services in our sample consist of a mix of the 100 most visited providers of the Alexa website ranking of Germany in September 2020 and personal accounts that the three authors hold. This selection ensures that the most important providers are covered by our analysis, but also that the performance of small providers is measured. In doing so, we have included any service from the Alexa top 100 website ranking, for which the following criteria hold true: first, registration is possible and for free; second, a language setting in either English or German is available; and third, the primary focus of the service is not on adult content. Where the authors did not already have a personal account at a provider with these criteria, one of them opened up a new account and created content over a couple of days, before requesting their personal data. Each author who sent a data export request examined the incoming dataset for valid information.

Over the same time period, 190 online services (including all 182 services in the data export sample and 8 additional services where a costly subscription would have been required to analyze the export) were examined on their data import capabilities. For each online service we collected data in three ways. Firstly, we logged into our existing or newly created accounts, accessed all available parts of the user interface and checked for import possibilities. Then, to see if there is any import possibility not directly offered via the user interface, we searched the service's documentation, i.e. help sections, FAQs, and similar, for mentions of ways to import or upload data. Finally, we used a search engine to search for combinations of $nameofservice and {import, upload, migration, data portability} and analyzed the first five results of each search query.

Lastly, we gathered NACE (Rev. 2) industry codes [13] from the Orbis database [30] for the online services in our sample. As the Orbis database contains more than 375 million companies' corporate information, we were able to find the required information for 179 of the online services in our sample.

3.2 Measures

To be able to conduct the formal analysis and ensure objectivity and comparability, we created high-level attributes that measure the data scope and the compliance (with the requirements of GDPR Art. 20). The scope is based on the taxonomy by De Hert et al. [9]; for the compliance we assessed the data formats using the compliance table by Wong and Henderson [11]. Table 2 gives an overview of all measures and their possible manifestations.

Data scope and completeness

For the analysis of the scope of personal data provided, we followed the data taxonomy by De Hert et al. [9]. We therefore inspected each service's data export and categorized the contained data (e.g., personal information, messages, or location data). In alignment with the examples provided in the *Guidelines on the Right to Data Portability* [29], we then classified the data as *provided*, *observed*, or *inferred*. See Appendix B for an overview of typical data exported per industry sector and their respective classifications.

Table 2. List of measures.

Measure	Scale	Range	Description
Format Compliance	binary	{true, false}	Format of the data export is compliant with the provisions of Art. 20 GDPR (structured, common, machine-readable) and data was provided within the legal time frame.
Overall Compliance	binary	{true, false}	Data export is format compliant and contains the legally required minimal data scope of received, i.e. actively provided, data.
Export Scope	ordinal	{No personal information available, received data, received & observed data, received & observed & inferred data}	Richness of the data provided in the data export (based on taxonomy of [9]).
Import Scope	ordinal	{None, minimal, partial, full}	Proportion of functionalities for which import possibilities are offered.
Authentication Factors	ratio	$\mathbb{N}_0$	Number of authentication factors that need to be provided to request the export and access the personal data.
Duration	ratio	$\{0, 1, \ldots, 90\}$	Days until Art. 20 request is completed.
Alexa Rank	ordinal	$\mathbb{N}$	Position in the Alexa page ranking. Services with a higher rank are more popular (rank x is higher than rank $y \Leftrightarrow x < y$)
Industry Sector	nominal	10 sectors	Industry classification based on NACE Rev. 2.

Note that [29] prompts providers to interpret the provision of personal data under GDPR Art. 20 broadly, i.e. including the category of *observed data*. However, in our analysis we only used *received data* to assess whether the data export was complete (overall compliance). We did this for two reasons: firstly, we can only assess the completeness of data that we, as authors, consciously provided when using a service, and cannot know how much observed data a service provider ultimately tracks about its users; secondly, the category of received data is the only one that online services apparently must provide to data subjects under the binding data portability GDPR legislation (in 2020).

Format compliance

To analyze the compliance of the data export with the provisions of the GDPR, we also considered the data format of the export.

We used the RtDP file format compliance table of Wong and Henderson [11] (see Table 1) to evaluate whether the exported data is structured and machine-readable. Formats where Wong and Henderson [11] found ambiguities (e.g., PDF and HTML) were treated as non-compliant. If the personal data was sent in more than one data format by the provider, we decided on one data format representing the most applicable one. In cases where both raw data and metadata were provided, we chose the metadata format. With this mechanism we would, for instance, select *the json-format* as the main format provided by Instagram because Instagram's most important provided data are the pictures/stories, which are either sent in a *jpg-format* or *json-format* by the provider.

Import scope

Regarding data import, we categorized the online services according to the import possibilities they offer. We therefore identified the *core* functionalities of each service (e.g., search, email, or banking) and examined whether data can be imported for any of these functionalities. Online services offering an option for importing all their core functionalities were categorized as *services which offer* ***full*** *import possibilities*, whereas those offering import for at least one but not all of their core functionalities were categorized as *services which offer* ***partial*** *import possibilities*. Services only offering import possibilities for minor functionalities but not for *core* functionalities (e.g., a survey service which offers to import contacts but not surveys themselves), were categorized as *services which offer* ***minimal*** *import possibilities*. All other services were categorized as *services which offer* ***no*** *import possibilities*.

Industry sectors

In order to compare data portability practices across industries, we categorized the services. Unlike [11], we

did not use the industry categorization according to curlie.org, as we considered its categorization ambiguous in many cases. We decided to create our own categories on the basis of the NACE Rev. 2 classification of Eurostat [13] as we realized that all official industry classifications such as NAICS or NACE do not adequately reflect today's online industries.

Therefore, we proceeded to establishing a limited number of industry categories by grouping companies with similar NACE codes. To ensure the highest possible objectivity, three authors and two persons unfamiliar with the topic first grouped the services independently of each other into 8 to 12 categories. After mutual disclosure of these categories, we agreed on 10 categories. In the second phase, we independently assigned all corporate NACE codes, which occur at least once in our sample, to the established industry categories. After disclosing individual allocations of NACE codes, we agreed on the final allocation of NACE codes to categories, as illustrated in Appendix E. Lastly, we examined whether the corporations fitted into the industry categories on the basis of their NACE codes. In a few unambiguous misallocations as well as for the 11 online services without an available NACE code, we assigned the corporations to what we considered to be an appropriate match.[3]

3.3 Hypotheses development

Although the GDPR has now been in force for more than two years - as of November 2020 - the "CMS.Law GDPR Enforcement Tracker" does not list a single fine issued by regulators to an online service for an Art. 20 violation [31]. As Wong and Henderson [11] have shown that not all services fulfill their legal duty, we assume that Data Protection Agencies have been reluctant to enforce data portability legislation - probably to give corporations sufficient time to adapt. However, for privacy legislation to be effective, legislators will likely increase the enforcement of the regulation in the coming years. We can already see an increase of the number of penalties for GDPR violations as the GDPR penalty enforcement tracker shows: whereas in 2019 the tracker counted 164 penalized violations across members of the European Economic Area, the number increased to 262 penalized violations in 2020 [31].

Therefore, we expect popular services that are exposed to the public to be especially concerned about being compliant with data portability regulation as penalty enforcement rises. We defined the online services' popularity by the proxy of their Alexa rank[4] [32]. We are interested in two interpretations of compliance. Firstly, in a base hypothesis, we want to examine compliance with regard to file format and valid time horizon (format compliance). We thus state:

H1a: Online services with higher popularity among consumers will be significantly more often *file-format compliant* with data portability legislation than online services with less popularity.

Secondly, we defined the *overall compliance* with GDPR Art. 20 in our study in a narrower way than the law or Wong and Henderson [11] do, as we included the demanded data scope as a relevant criterion. Thus, for a service to be compliant, it would need to have sent the personal data in a compliant file format, within the valid time horizon, and providing a complete set of any data the subject has actively provided ("received" category of De Hert et al. [9]), whenever it made use of the service in the past two years since the enactment of the GDPR. We thus state:

H1b: Online services with higher popularity among consumers will be significantly more often *overall compliant* with data portability legislation than online services with less popularity.

Following theoretical analyses on data portability and the economics of data in the literature [1, 4], we would furthermore advance that popular online services with a high Alexa rank have a high incentive to give out only the data scope that the regulation requires from them. This is based on the following reasoning: as services with higher ranks are frequently market-leading in their sectors, they control the data collection for a key service in this industry [2]. Google, for example, controls the data collection on search queries. If services with higher ranks want to protect their market-leading position towards competitors or new entrants, they strategically should comply with the regulation, but with the smallest data scope possible in order to

3 Readers can request the sample with corporate names, NACE codes, and final group allocations for inspection from the authors.

4 In the regression analysis, we were obliged to use the worldwide Alexa ranks, since for 65 online services in our sample a German rank was not available. We could find the worldwide Alexa ranks for all services in our sample.

make it harder for data subjects to switch. If a person, for example, wants to switch their music streaming application, but can only bring their actively-provided data to the new streaming application, they will not perceive the service as positively in the new application as they had perceived it in the old application because the new provider lacks observational data on their streaming history. Observational data are especially useful since the new provider can process them to infer listening interests. We thus state:

H2: Online services with higher popularity among consumers will not provide data subjects with more data scope than online services with less popularity do.

Apart from the scope, services must ensure that the personal data is transferred to the correct person to prevent data leaks and possibly legal penalties. Therefore, all services should have strong authentication methods in place, in the best case scenario, multiple authentication factors [33]. It can probably be assumed that popular online services, which carry a higher responsibility, since more consumers interact via it, possess more human and financial resources to deploy and manage a higher number of authentication factors than their competitors or new entrants. We thus expect:

H3a: Online services with higher popularity among consumers will use more authentication factors to identify individuals correctly.

In a similar vein, services that provide data subjects with a large data scope should have worked on authentication more thoroughly than services that do not actively, but rather passively, manage data portability regulation. We thus expect:

H3b: Online services which provide data subjects with a larger data scope will use more authentication factors to identify individuals correctly.

Next, a shorter time duration serves to lower switching costs of consumers between online service providers. Therefore, it would be interesting to see whether more popular platforms in general provide their data exports faster than their competitors. We would expect that popular platforms with many users have more financial resources to automate the data export process. Additionally, smaller platforms probably receive fewer requests that they decide to handle manually in order to not incur the initial setup costs of automating data export requests. We thus state:

H4: Online services with higher popularity among consumers will provide the transfer of data significantly faster than their competitors.

One of the core motivations behind the RtDP is to foster competition between platforms and to facilitate market entry for new platforms. In his theoretical model, Wohlfarth [4] shows that the implementation of a RtDP yields higher expected gains in profits for market entrants than for incumbent services. Therefore, we expect that less popular platforms (i.e., platforms with a lower Alexa rank) would offer more extensive ways to import data from other platforms in order to increase their market shares.

H5: Online services with lower popularity among consumers will provide more and better import opportunities for data subjects than their competitors.

4 Results

4.1 Descriptive statistics

Regarding the descriptive findings, we first outline the analysis of the data exports, followed by the data imports. Where applicable, we compare our results to those gathered by Wong and Henderson [11], who conducted their study in 2018 immediately after the GDPR came into effect.

Requesting the data export: duration, authentication, and transmission

As described in Section 3.1, we analyzed the data export of 182 online services. Of these services, 68 (37.3%) offer a predefined way for requesting data exports under GDPR Art. 20: At 45 service providers we could issue a request via a simple button click within the service portal itself, whereas 23 services providers offered an online request form in the privacy sections of their websites.

Only 135 of the 182 services (74.2%) managed to execute some sort of data export in the legally valid time horizon, which is similar to the 74.8% observed by Wong and Henderson [11]. 24 services failed, for instance due to originally confirming our request but then never fulfilling it, and another 23 services did not even respond in the first place. Of these non-executed data exports, more than 70% have been requested via email.

The following evaluations will, if not stated otherwise, only consider the 135 services that executed a data export. The RtDP demands the data export to be executed within 30 days, unless a service provider asks

for an extra extension of another 60 days. In our study, six services requested an extension. On average, the duration of the data export was 9.5 days with a median execution time of 4 days. 38 requests were even fulfilled on the same day. Thus, the response time has decreased significantly compared to the study of Wong and Henderson [11] who observed a median time of 19 days. Moreover, we noticed that the execution time varies considerably between industries. The median time for services in the *Hardware & Software Manufacturers* industry is 0 days, and 0.5 days in the *Social Communities & Messaging* industry. Much slower responses were observed in the *Financial & Legal Services* industry were the median execution time was 16.5 days.

Regarding authentication, Di Martino et al. [34] observe that data controllers have widely different policies in place for "Right of Access" requests (GDPR Art. 15) and that some of them are even prone to social engineering attacks. While we did not specifically check for vulnerabilities, we also observed that the requirements which services placed upon us before sending out the data varied substantially. For our RtDP requests, only 104 services (77.0%) required some sort of authentication. The remaining 31 services required no authentication at all, which usually meant that they directly responded to our initial email with an email containing the requested data export.

The number of necessary authentication steps range from 0 to 3 with a mean of 1.0. When grouping the authenticating steps requiring knowledge (e.g., login data, proof of access to email account) to one factor, the number of services requiring a two-factor authentication was 7 (5.2%).

The services had individual ways of transmitting the personal data, which we arranged in four superordinate categories: mail (via postal service), email (including the exported data as an attachment or protected file), download link sent via email, and download option from the service directly (portal download or chat channel). The first category is the smallest with only 8 data export executions, email was used 44 times, download link via email 31 times, and last but most frequent, service download options with a total of 52 executions.

Graphs on duration, authentication, and transmission methods per industry sector are provided in Appendix C.

File formats, scope, and compliance

Table 3 lists all file formats that are present in at least 10 data exports.

Table 3. Most common file formats in data exports.

Format	Observations
CSV	36
JSON	30
PDF	14
XLS/XLSX	14
HTML	13

Out of the fully GDPR-compliant formats, identified by Wong and Henderson [11], CSV, JSON, and XML, which are also the recommended ones in the *Guidelines on the Right to Data Portability* [29], are detectable in our data exports. These format-compliant observations add up to a total of 69 online services (51.1%).

20 of the 135 services that executed the data export misinterpreted our RtDP request – which we specifically stated – as a request for RoA. One major difference in the RtDP is the need for the services to provide the data in a structured, commonly used and machine-readable format. This confusion seems to be one reason as to why some services failed to comply with these format requirements.

When analyzing the scope of the data according to the data taxonomy of De Hert et al. [9], we find that 132 services have exported received data, of which 84 also exported observed data, and of which 12 services

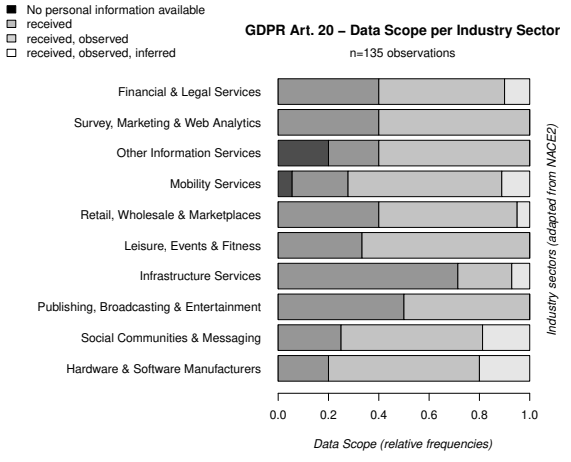

Fig. 1. Export scope per industry sector.

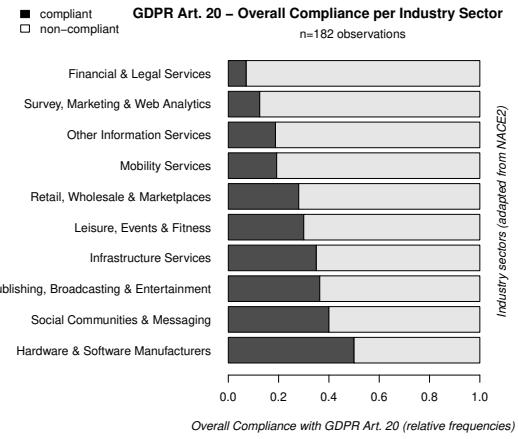

Fig. 2. GDPR overall compliance per industry sector.

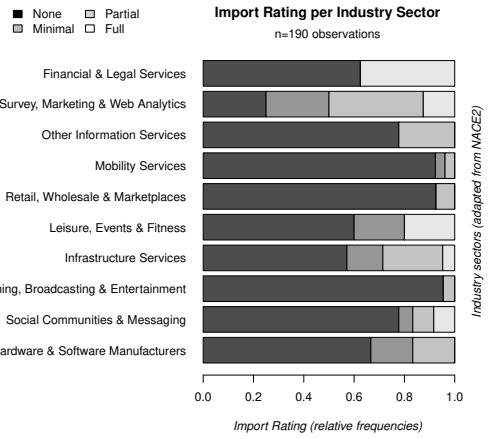

Fig. 3. Import scope per industry sector.

even exported inferred data. In general, we did not find evidence of incorrect information in the exported data. Besides, three services plausibly demonstrated that they do not store any personal data. Figure 1 shows the scope of the exported data per industry sector. In all industries except *Infrastructure Services*, more than 50% of services provide more data than required by the GDPR by including at least some observed data.

Regarding our completeness criterion, 101 of the 132 services have exported all of the personal data we have actively provided to them (i.e. received data category), so that we therefore categorized them to have conducted a *complete data portability export*.

In order for the personal data export to be overall compliant with the GDPR, it has to be executed within the given timeframe of 30 days (90 days if extension applies), be transmitted in a structured, commonly used, and machine-readable format, and be complete. From the original 182 data export requests sent only 52 services fulfill all of these criteria. Therefore, 130 services (71.4%) failed in some way to comply with the request on RtDP.

The GDPR overall compliance per industry sector is apparent from Figure 2. In none of the industry sectors are more than 50% of services compliant with the RtDP. The lowest share of GDPR-compliant services can be found in the *Financial & Legal Services* industry where more than 90% of services in our sample are not compliant, primarily due to the choice of inadequate data formats.

Importing data

In addition to the data export, we also analyzed if and to what extent online services offer the possibility of importing data. We found that none of the analyzed 190 services directly offer to import data generated by a RtDP request, e.g., data exported from Facebook. Using the categorization by core functionalities, described in Section 3.2, 146 services (76.8%) offer no import possibilities at all, while 44 (23.2%) offer at least minimal import possibilities. Of these 44 services, 11 offer minimal data import, 20 offer partial data import, and 13 offer full data import.

Figure 3 shows the import scopes grouped by industry. Two notable industries in this context are *Financial & Legal Services* and *Survey, Marketing & Web Analytics*. The financial services show the highest share of *full import* among all industries which is very likely a result of the Payment Accounts Directive 2014/92/EU [35] requiring banks to facilitate payment account switching. The services in the *Survey, Marketing & Web Analytics* industry have the lowest share of *no import* with 75% offering at least minimal import.

As for the 182 services for which we analyzed both export and import scopes, we were also interested in possible relationships. In Table 4, the relative frequencies of services which offer no import possibilities and services which offer at least minimal import possibilities are shown, grouped by export scope. It can be seen that services with a higher scope of exported data more often offer the possibility of importing data from other services.

Table 4. Import support by export scope.

	No import	Import
No personal info available	1.00	0.00
Received	0.81	0.19
Received, observed	0.76	0.24
Received, observed, inferred	0.67	0.33

4.2 Regression results

We used logit, ordered logit and OLS regressions to further examine whether certain properties of online services, such as their popularity with consumers as measured through the proxy of their Alexa ranks, exert significant influence on the compliance, usefulness and effectiveness of their data portability provisions.

Table 5 shows the results of the regressions. The LaTeX table was generated with the R stargazer package [36], ordered logistic regressions were computed with the R MASS package [37].

In Hypotheses 1a and 1b, we presumed that online services, which enjoy a higher popularity with consumers should be under more pressure to fully comply with data portability regulation than services that are visited by less consumers and less frequently. We therefore regressed the services' compliance capabilities on their Alexa ranks. As noted in more detail in the descriptive statistics section, a service is regarded as format-compliant if it successfully executes a RtDP request in the legally-allowed time duration and provides the format in a machine-readable and structured format. The overall compliance variable additionally includes the requirement that the service provided a complete dataset of the personal data that the data subject has actively provided to it. Moreover, similar to [38] we logarithmized the Alexa rank to better account for big numerical differences in ranks.

In Table 5, we can see that the results do not support Hypothesis 1b. They do not show that overall RtDP compliance is significantly higher for more popular platforms. Hypothesis 1a is weakly supported by the data in our sample. We find at a significance level of 10% that popular services with a higher Alexa rank comply significantly more often with the GDPR Art. 20 format requirements.

However, complying more often with privacy regulation does not necessarily imply that popular services aim to further improve the usefulness of data portability for consumers. From a strategic point of view, we expected in Hypothesis 2 these services not to provide a larger scope of personal data than their less popular competitors with a lower Alexa rank do, in order that they could maintain their collection privilege of certain personal data [2, 39]. As the siloing of certain personal data is one factor for protecting incumbents from competitors and new entrants, it would certainly be rational for a popular online service to take such an action.

In contrast to what might be expected from the literature, we find with our second regression that, when requested, online services with a higher Alexa rank, provide observed and partially even inferred data to individuals significantly more often. It therefore seems that contrary to our prediction, popular services demonstrate a willingness to comply with privacy regulation to a larger extent. A possible explanation for this result is that no direct infrastructure for data portability currently exists. Executing an indirect data transfer is more complicated and, due to a lack of import possibilities, often not feasible for a consumer. Therefore, popular services might not be affected by negative economic implications of providing a larger data scope. A second explanation for this result could also be that smaller services simply collect less observational data and infer less insights out of their data assets in contrast to popular services.

Although the current behavior of popular services to provide a comparatively large data scope serves to strengthen the notion of direct data transfers between online services, it is a risky strategy to provide a large scope of personal data portability in a privacy regulation regime such as the GDPR, when services do not employ strong authentication measures before exporting data. In Hypothesis 3a, we therefore firstly expected popular services, which stand in the spotlight of regulators, to require significantly more authentication factors for personal data export than their competitors. In Hypothesis 3b, we further assumed that the Alexa rank of a service might not necessarily be decisive in determining the number of authentication factors employed by it, but rather the provided data scope. Table 5 demonstrates that we find support for both hypotheses. In the light of the results of the second regression, this result makes sense, since popular services are simultaneously the services that more frequently provide the largest data scope.

In order to comprehensively investigate how well online services implement the requirements of data portability regulation, we also monitored the time that services took to fulfill the data portability request of a subject. We could already show that on average, providers needed less time in our sample in comparison to the

Table 5. Effect of popularity on GDPR compliance, data scope, authentication, and duration.

	Dependent variable:						
	Format Compliance	**Overall Compliance**	**Export Scope**	**Authentication Factors**		**Duration**	**Import Scope**
	logistic		*ordered logistic*	*ordered logistic*		*OLS*	*ordered logistic*
	(1a)	**(1b)**	**(2)**	**(3a)**	**(3b)**	**(4)**	**(5)**
log(Alexa)	**−0.106***	**−0.048**	**−0.161****	**−0.225*****	**−0.201****	**−0.248**	**−0.221*****
	(0.061)	**(0.063)**	**(0.065)**	**(0.078)**	**(0.080)**	**(0.387)**	**(0.075)**
Export Scope					**0.738****		
					(0.334)		
Industry							
Social Communities & Messaging	−0.219	−0.323	0.240	0.231	0.200	−2.679	−0.011
	(0.944)	(0.895)	(1.015)	(1.210)	(1.227)	(5.929)	(0.931)
Publishing, Broadc. & Entertainment	−0.964	−0.426	−1.006	−1.265	−1.061	4.155	−1.511
	(0.992)	(0.948)	(1.105)	(1.294)	(1.318)	(6.623)	(1.340)
Infrastructure Services	−0.453	−0.650	−1.411	0.463	0.792	−1.841	1.357
	(1.014)	(0.988)	(1.122)	(1.349)	(1.366)	(6.587)	(0.986)
Leisure, Events & Fitness	−0.997	−0.598	−0.026	−1.224	−1.264	3.593	1.546
	(1.155)	(1.122)	(1.191)	(1.390)	(1.414)	(7.222)	(1.124)
Retail, Wholesale & Marketplaces	−1.162	−0.812	−0.551	−0.322	−0.169	2.601	−0.990
	(0.985)	(0.948)	(1.052)	(1.258)	(1.276)	(6.239)	(1.134)
Mobility Services	−1.290	−0.926	0.432	0.263	0.209	12.635*	−0.388
	(1.044)	(1.012)	(1.135)	(1.342)	(1.358)	(6.655)	(1.185)
Other Information Services	−0.908	−1.331	−1.178	−0.492	−0.129	6.311	0.361
	(1.028)	(1.055)	(1.153)	(1.346)	(1.383)	(6.750)	(1.039)
Survey, Marketing & Web Analytics	−0.818	−1.771	−0.602	−1.099	−0.982	13.620*	2.375**
	(1.160)	(1.366)	(1.292)	(1.488)	(1.498)	(7.852)	(1.084)
Financial & Legal Services	−1.946	−2.318*	−0.123	1.905	1.993	9.805	2.150*
	(1.198)	(1.362)	(1.175)	(1.451)	(1.469)	(7.011)	(1.104)
Constant / Intercept 1	1.153	0.197	−5.645***	−3.244***	−1.040	7.920	0.042
	(0.914)	(0.861)	(1.151)	(1.197)	(1.551)	(5.623)	(0.871)
Intercept 2			−2.161**	1.571	3.968**		0.484
			(0.983)	(1.153)	(1.611)		(0.870)
Intercept 3			0.976				1.679*
			(0.964)				(0.894)
Observations	182	182	135	135	135	135	190
R^2 / *Nagelk. Pseudo* R^2	*0.128*	*0.084*	*0.154*	*0.215*	*0.255*	0.174	*0.243*
Adjusted R^2						0.108	
Log Likelihood	-111.814	-103.395	-125.732	-86.201	-83.704		-129.394
Akaike Inf. Crit.	245.629	228.790	277.464	196.401	167.408		284.788
Residual Std. Error						12.156 (124)	
F Statistic						2.621*** (10; 124)	

Notes. The table reports the effect of popularity on different characteristics of data transfers under Art. 20 GDPR. Standard errors are in parentheses below the estimates. *p<0.1; **p<0.05; ***p<0.01

Alexa rank is used as a proxy for popularity (higher rank implies higher popularity, x higher than $y \Leftrightarrow x < y$), logarithmized to account for positive skewness. Dummy variables for the industries are used as control variables. Format compliance (1a) indicates whether format and duration of the data export are compliant with the provisions of the GDPR, overall compliance (1b) additionally takes the completeness of the received data into account. Export scope (2) takes values from 1 ("no personal info available") to 4 ("received, observed, and inferred data") and indicates how rich the scope of the data export is according to the taxonomy of [9]. Authentication factors (3a, 3b) describes how many factors a person needed to provide to request or access the data export. Duration (4) describes the number of days a service needed to process the Art. 20 request. Import Scope (5) takes values from 1 ("no import") to 4 ("full import") and describes to which extent import possibilities are offered.

sample of Wong and Henderson [11]. Focusing exclusively on our sample, we assumed with Hypothesis 4 that more popular online services are faster in providing the data export. However, our results do not support this hypothesis. Table 5 shows no significant impact of a service's Alexa rank on the speed of data export.

Lastly, we supposed that smaller online services and entrants will offer consumers more import possibilities. In contrast with this theoretically motivated expectation, we find that more popular platforms currently provide more import possibilities than their competitors. This effect is highly significant (at 1% level)[5]. The evidence of our sample therefore contradicts the assumption of the model of Wohlfarth [4] that small platforms and entrants currently think strategically about the chances that data portability regulation can provide to them. This might however change with time.

All in all, the regression results support the observation that popular online services seem to have less fear about engaging in data portability than one might have expected before our study. Notwithstanding, it is important to not expand consideration of the result to beyond the current situation, in which an infrastructure for direct data portability between providers is not yet established. As soon as such an infrastructure exists, it remains likely that popular services will adapt and provide a smaller data scope to defend their data collection privilege and maintain their position as indisputable market leaders.

5 Discussion

5.1 Addressing the data portability divide and consumer indifference

Our results show that for a majority of online services data portability currently remains a burden placed upon them by the legislator. They comply with this burden, but do not make use of data portability as originally intended by policy makers and academic scholars [8]. In contrast to what could be expected in theory, we do not find evidence that entrants and competitors to market-leading services try to profit from the potential to receive a free-of-charge data copy via consumers, which could fuel their businesses and increase service qualities to a level at which they would be able to compete with incumbents [40].

Take a maps service as an example. A majority of consumers uses a maps application every day for navigating their vehicle, finding new bars and restaurants, obtaining orientation in a new city or saving their favorite leisure time locations. As we know from [40], online services, such as maps services, provide their greatest utility to users the more received and observed data they are able to collect. Users will prefer to use the maps application that shows them closed streets upfront, has valid opening times of shops and can calculate where traffic jams will most likely occur. However, in our sample, maps services do not even offer users the option of importing data.

As with maps services, online services in many industries currently do not offer or only offer limited means for importing data. Some industries, however, represent an exception to this rule as the examples of the financial and the fitness industries shows. In both industries almost all online services provide the opportunity to comprehensively import data. In the case of the finance industry, this development was triggered via the Payment Accounts Directive 2014/92/EU and Payment Services Directive 2 [41]. In the case of the fitness tracking industry, we are not aware that a regulation was needed, but services seem to have started to offer import capabilities due to their interest in the fitness data of individuals. We therefore observe that there is no single way that leads towards more corporate or industry awareness of data portability as an opportunity for growth, but rather that several paths exist.

Table 6. Means of export and import scope ratings (1: worst, 4: best) by Alexa rank.

	Export	Import
Alexa $<$ 400	2.93	1.58
400 $\leq$ Alexa $<$ 6500	2.56	1.51
6500 $\leq$ Alexa	2.57	1.33

Moreover, our sample reveals that currently the most popular incumbent services provide the largest data scope upon a consumer's data export request whereas market entrants such as smaller services are rather restrictive in what they export (see Table 6). This partially contradicts what the strategy and economics literature would have expected from the introduction

5 This result also holds true when the import scope categories *minimal* and *partial* are merged, i.e. when the distinction between *core* and *non-core* functionalities is omitted. The corresponding regression table is provided in Appendix D.

of a data portability regulation: while a new entrant or service competitor (with small market share and few consumers or data to lose) could provide a large data export scope and many options for easy data import to gain consumer trust, we find on average that these services are rather limited in both domains. By contrast, an incumbent has little interest in providing a large export scope, since it could easily lose consumers and data to competitors. However, we find these online services to be the most generous in terms of the exported data scope and possibilities to import data.

Thus, we currently seem to observe a *data portability divide*, in which a few large incumbents strategically make use of the advantages that data portability can bring to them, such as consumer trust, which might lead to higher data provision in the first place [26]. Lam and Liu [26] have called this phenomenon the *demand expansion effect*. Similarly, they mitigate the risks that a more effective data portability regulation could place upon them by lobbying for less regulation [9].

On the other side of the divide, a majority of corporations try to comply with a regulation whose economic implications they do not seem to grasp or at least do not manage to use for their own growth and capability to innovate.

As with the majority of corporate actors, many consumers seem to be unaware of or do not appreciate the fact that data portability regulation preserves their freedom to not become dependent on single online service providers in the future [42]. Future research is necessary to distinguish whether the current obstacle of not having a direct possibility for transferring data between providers or a lack of interest hinders them from engaging more with data portability. If the latter proves to be the case, then policy makers need to reconsider their assumption that a majority of the population is interested in preserving a right that requires providers to keep switching costs for consumers low, while data portability implementation and maintenance costs for online services rise.

5.2 Towards direct data portability between providers

In our study, we have observed that an indirect transfer of data from one provider to another (we call it *indirect data portability*) is currently only rarely possible as the share of online services offering some sort of data import is rather low (23.2%). A possible explanation for the low share of services that make use of the advantages of the RtDP is that developing mechanisms for importing data can come with a considerable cost [25]. Furthermore, a supported service could change the structure of exported data at any time, requiring a redevelopment of the respective import mechanism.

In contrast to *indirect data portability*, which we have examined with this study, a direct transfer between data providers can take on various forms. Application programming interface (API) adapters can, for example, be built between two or multiple providers, who wish to exchange data.

Currently, the most common way for two providers to connect their services is for one incumbent service with the privilege of data collection in a certain area to export data to a smaller service. This data exchange, which functions frequently via the OAuth authorization protocol [43, 44], is however a one-way data exchange by design. In contrast to data export requests using the RtDP, where services are obliged to export data, it is a service's choice as to whether it offers a data transfer using OAuth and about which data is shared.

Table 7. Most common OAuth data providers.

Data provider	Observations
Facebook	69
Google	60
Apple	16
Twitter	7
Verimi	5
Microsoft	4
LinkedIn	4

We found that 45% of online services in our sample support data import using OAuth, usually via a separate login button. The supported number of data providers was between 0 and 8, with a mean of 0.99. Table 7 shows the data providers that were most commonly supported. It can be seen that Facebook and Google with 69 and 60 observations, respectively, are the predominant data providers.

These results show that there is indeed a demand for data portability for a large number of online services.

6 Avenues for future work

Data transfers using OAuth can serve as an example on how direct data portability can be implemented. It is

easy to use, can be implemented with moderate effort, and data is transferred immediately upon request. However, it is not scalable without considerable effort, for example were consumers to demand that an online service offered import options from a multiplicity of providers.

Therefore, the second way to implement data portability takes a more universal approach. It does not aim to connect service APIs on a one-to-one provider basis. Instead, it targets establishing a platform to which each online service only needs to connect once in order to be able to exchange data with all services connected up to that point in time and in reciprocal ways (see, for example, Data Transfer Project [14] or Data Portability Cooperation [45]).

For both manifestations of data portability (OAuth, multiple-interface platform), we see a couple of academic and practical challenges that still have to be overcome before the RtDP's postulated vision of a more economy-spanning direct data portability infrastructure can become a reality:

First, in order to know what effort is necessary to ensure individuals' future right to easily switch between service providers, more empirical studies on data portability at the individual level are needed. We need to better understand in which cases of particular data settings a majority of individuals would feel a real need to transfer their past personal data. For instance, we know that people value porting their social network history [46], but for many data settings we do not have such clear evidence or only evidence for certain geographical regions [42]. We therefore encourage researchers to elaborate on these initial results and cross-validate them for various regions.

A second challenge in making data portability more attractive to consumers is posed by the question of how to encourage more corporations to increase their data import options. This could, for instance, be achieved by convincing them to allow for more OAuth data providers or by encouraging them to commit to building an adapter to a multiple-interface data portability platform. Although the latter problem is related to the well-known "chicken-and-egg" problem of how to get a platform started [47], the situation is slightly different in the case of a data portability platform. Data portability is a legal obligation and, thus, corporations want to minimize costs to comply with it.

Therefore, we envisage three ways in which to make progress: firstly, startup founders and small business operators could be educated on the advantages RtDP offers for market entrants or any firm that is not a market leader in order to popularize the integration of more import options. Secondly, besides educating, regulators could consider incentivizing firms to join a multiple-interface data portability platform by granting some kind of benefit, e.g., a reduction in taxes, in exchange. As an alternative to encouraging voluntary action, regulators could also mandate corporations of certain industries, in which consumers verifiably reported the desire to transfer their data in more easy ways, to build an adapter. One solution could be to find indicators for each industry on when such an obligation is realistically feasible. Possible indicators include, for instance, the number of users or the market share of a service necessary for ensuring that no regulatory overburdening negatively impacts new market entrants or small platforms [25].

Thirdly, our export data reveal a considerable fragmentation in data formats and richness. The data format results highlight that many standard setting procedures are necessary to ultimately enable a smooth direct data portability implementation as for example with a multiple-interface data portability platform. For instance, data models in certain data settings need to be defined, so that different data formats can be adequately transformed from the format of the exporting online service to the format of the importing service [14]. For instance, one fitness service might save its data in a .tcx format, while another one uses the .hrm format.

In addition, further academic progress on the topic could lie in a continued dive into the richness of exported personal data. We examined the scope of exported data as one measure for data richness. Yet, building on our analysis, another empirical study at the organizational level could assess the *rate of reusability* of exported data. What fraction of received, observed or inferred data can actually currently be reused by other platforms? What fraction could in theory be reused, but is currently not importable at most services?

Regarding authentication, we found that services with higher Alexa rank required more proofs of authentication and that the types of authentication methods used still vary significantly within each industry (compare Appendix C). Adding to our results with a first-party sample, Urban et al. [48], for example, showed that in some cases third-party tracking companies raised unreasonable access hurdles to the transfer of personal data on the basis of GDPR Art. 20 by the requirement of signed affidavits or copies of official ID documents for authentication. To ease the data export request process for consumers, we therefore believe that authentication methods for data portability should be standardized and based on the sensitivity of requested data.

The process of standardization should be accompanied by further empirical studies to measure whether and how quickly services adopt standards and to evaluate whether further regulatory steps are necessary.

7 Conclusion

More than two years ago, European legislators established a Right to Data Portability (RtDP), which should allow consumers to transfer their data directly from one online service to another. In theory, the right was established to strengthen individuals' control over their personal data, to lower switching costs of consumers between providers, and to ultimately act as one possible remedy against increasing tendencies of winner-take-all scenarios in online markets. Yet, in 2020 the RtDP still does not live up to the vision of its originators in practice, as an infrastructure to enable direct data transfers between providers still remains to be finalized.

Under these circumstances, we have empirically examined how easy an individual with interest in pursuing his right can actually transfer data between online services in an indirect way. We define such indirect porting as a two-step transfer: firstly, requesting a data export from one provider, followed secondly by the import of the obtained data at another provider. To answer our research question we requested data exports from 182 online services, which include the Alexa top 100 websites in Germany as of September 2020, and simultaneously inspected their data import capabilities.

We find that the popularity of online services, as measured by their Alexa rank, does not only influence the compliance with data portability legislation or the number of authentication factors required to verify the requester's identity. It also exerts a significant positive influence on the data scope that services export to individuals and the import possibilities that they offer to them.

The latter results are quite surprising and in contrast to what game-theoretic literature on data portability would expect. In theory, service entrants or small competitors were expected to be in a better position to challenge market dominating online services with the support of the RtDP, as they could offer larger export scopes and import capabilities than incumbents, which might fear opening up their data silos and losing consumers. Yet, our sample shows that currently, incumbents are most generous in exporting data and offering opportunities to import data, which leads to the assumption that their smaller competitors have not yet fully grasped the original intention of the legislator's regulation design. Incumbents, however, seem to know better how to use the RtDP for defending their positions by building enhanced trust with consumers, which in turn can lead to them providing more data.

In addition to these main results, we also clustered online services into industry sectors using the NACE Rev. 2 industry classification in order to investigate possible differences in progress on data portability compliance across industries. As a result, we observe that especially financial services (as a result of the Payment accounts directive 2014/92/EU) and fitness services (not driven by sector-specific legislation) provide individuals with extensive import possibilities. We thus conclude that improved data portability implementation can be fostered by both, stricter regulation or industry self-commitment.

As with any study exploring new research fields, our study is subject to limitations. Firstly, our results should only be generalized with caution until our results are replicated with larger corporate samples that reduce potential biases from convenience sampling. Secondly, it is important to note that our sample deviates from the sample of Wong and Henderson [11], as our sample includes more online services used by German citizens. And lastly, we are aware of potential subjectivity bias in our industry categorization. Therefore, we aimed to categorize online services as objectively as possible by building on NACE codes. Our study also highlights that current industry categorizations barely reflect the diversity of online markets that have developed over the past decade. We therefore call upon institutions to renew their industry categorizations.

Ultimately, our research aims to inform regulators as well as small online service operators alike. The former would be well advised to better educate digital entrepreneurs and small service operators about the opportunities for growth and innovation that a well-implemented data portability infrastructure and legislation would offer to them. Moreover, we identify four main challenges and suggest pathways on how the current status of data portability between providers in practice could be improved and adapted so that it becomes more effective in the future. Lastly, our results should encourage digital entrepreneurs and small service operators to implement more data import possibilities so that individuals who are in search of alternatives to mainstream incumbent services can smoothly transfer to them without being turned away.

Acknowledgments

We would like to thank Martin Degeling, Katharina Hartinger, Carmen Loefflad, Robert Lusza, and the anonymous reviewers for their helpful feedback. We further are grateful for funding support from the Bavarian Research Institute for Digital Transformation (bidt). Responsibility for the contents of this publication rests with the authors.

References

[1] C. Jones and C. Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858, 2020.

[2] J. Krämer, P. Senellart, and A. de Streel. Making data portability more effective for the digital economy. *CERRE Report*, 2020.

[3] I. Graef. Mandating portability and interoperability in online social networks: Regulatory and competition law issues in the European Union. *Telecommunications Policy*, 39(6): 502–514, 2015.

[4] M. Wohlfarth. Data portability on the internet. *Business & Information Systems Engineering*, 61(5):551–574, 2019.

[5] A. Sunyaev, N. Kannengießer, R. Beck, H. Treiblmaier, M. Lacity, J. Kranz, G. Fridgen, U. Spankowski, and A. Luckow. Token economy. *Business & Information Systems Engineering*, pages 1–22, 2021.

[6] S. Mager and J. Kranz. Stimulating economic growth by unlocking the nonrival potential of data - Review, synthesis and directions for future research. *Available at SSRN 3720114*, 2020.

[7] S. Zuboff. Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1):75–89, 2015.

[8] I. Graef, J. Verschakelen, and P. Valcke. Putting the right to data portability into a competition law perspective. *Law: The Journal of the Higher School of Economics*, pages 53–63, 2013.

[9] P. De Hert, V. Papakonstantinou, G. Malgieri, L. Beslay, and I. Sanchez. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, 34(2):193–203, 2018.

[10] T. Linden, R. Khandelwal, H. Harkous, and K. Fawaz. The privacy policy landscape after the GDPR. *Proceedings on Privacy Enhancing Technologies*, 2020(1):47–64, 2020.

[11] J. Wong and T. Henderson. The right to data portability in practice: Exploring the implications of the technologically neutral GDPR. *International Data Privacy Law*, 9(3):173–191, 2019.

[12] S. Turner, J. Galindo Quintero, S. Turner, J. Lis, and L. M. Tanczer. The exercisability of the right to data portability in the emerging Internet of Things (IoT) environment. *New Media & Society*, 2020.

[13] Eurostat. NACE rev. 2 - Statistical classification of economic activities in the European community, 2008. Available at: https://ec.europa.eu/eurostat/web/nace-rev2/overview.

[14] B. Willard, J. Chavez, G. Fair, K. Levine, A. Lange, and J. Dickerson. Data Transfer Project: From theory to practice, 2018. Available at: https://services.google.com/fh/files/blogs/data-transfer-project-google-whitepaper-v4.pdf.

[15] I. Graef, M. Husovec, and N. Purtova. Data portability and data control: Lessons for an emerging concept in EU law. *German Law Journal*, 19(6):1359–1398, 2018.

[16] P. Klemperer. Markets with consumer switching costs. *The Quarterly Journal of Economics*, 102(2):375–394, 1987.

[17] DLA Piper. Data protection laws of the world, 2020. Available at: https://www.dlapiperdataprotection.com/system/modules/za.co.heliosdesign.dla.lotw.data_protection/functions/handbook.pdf?country=all, last accessed: November 25th, 2020.

[18] M. E. Porter. The five competitive forces that shape strategy. *Harvard Business Review*, 86(1):25–40, 2008.

[19] C. J. Hoofnagle and J. King. Consumer information sharing: Where the sun still don't shine. *Available at SSRN 1137990*, 2007.

[20] L. Thomas and C. J. Hoofnagle. Exploring information sharing through California's 'Shine the Light' law. *Available at SSRN 1448365*, 2009.

[21] S. Grogan and A. M. McDonald. Access denied! contrasting data access in the united states and ireland. *Proceedings on Privacy Enhancing Technologies*, 2016(3):191–211, 2016.

[22] J. L. Kröger, J. Lindemann, and D. Herrmann. How do app vendors respond to subject access requests? a longitudinal privacy study on ios and android apps. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, 2020.

[23] J. Krämer. Personal data portability in the platform economy: Economic implications and policy recommendations. *Journal of Competition Law & Economics*, 2020.

[24] J. Krämer and N. Stüdlein. Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the General Data Protection Regulation. *Economics Letters*, 181:99–103, 2019.

[25] P. Swire and Y. Lagos. Why the right to data portability likely reduces consumer welfare: Antitrust and privacy critique. *Maryland Law Review*, 72(2):335–380, 2013.

[26] W. Lam and X. Liu. Does data portability facilitate entry? *International Journal of Industrial Organization*, 69:102564, 2020.

[27] Council of European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04.

[28] G. Malgieri. Property and (intellectual) ownership of consumers' information: A new taxonomy for personal data. *Privacy in Germany - PinG*, 2016(4):133, 2016.

[29] Article 29 Data Protection Working Party. Guidelines on the right to data portability, 2017. Available at: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233.

[30] Bureau van Dijk. Orbis, 2020. Available at: https://orbis.bvdinfo.com/, last accessed: October 15th, 2020.

[31] CMS. GDPR enforcement tracker, 2020. Available at: https://www.enforcementtracker.com/, last accessed: November 26th, 2020.

[32] Alexa. Topsites in Germany - Ranking, 2020. Available at: https://www.alexa.com/topsites/countries/DE, last accessed: September 30th, 2020.

[33] M. Bishop. *The art and science of computer security*. Addison-Wesley Longman Publishing, 2002.

[34] M. Di Martino, P. Robyns, W. Weyts, P. Quax, W. Lamotte, and K. Andries. Personal information leakage by abusing the GDPR 'right of access'. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS)*, 2019.

[35] The European Parliament and the Council of the European Union. Directive 2014/92/EU. *Official Journal of the European Union*, 2014. Available at: https://eur-lex.europa.eu/eli/dir/2014/92/oj.

[36] H. Hlavac. *stargazer: Well-formatted regression and summary statistics tables*, 2018. Available at: https://CRAN.R-project.org/package=stargazer.

[37] W. Venables and B. Ripley. *Modern applied statistics with S*. Springer-Verlag, fourth edition, 2002.

[38] M. Zhao, J. Grossklags, and P. Liu. An empirical study of web vulnerability discovery ecosystems. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 1105–1117, 2015.

[39] S. Mager and J. Kranz. Consent notices and the willingness-to-sell observational data: Evidence from user reactions in the field. *Proceedings on the European Conference on Information Systems (ECIS)*, 2021.

[40] C. Argenton and J. Pruefer. Search engine competition with network externalities. *Journal of Competition Law and Economics*, 8(1):73–105, 2012.

[41] P. Constantinides, O. Henfridsson, and G. Parker. Introduction - Platforms and infrastructures in the digital age. *Information Systems Research*, 29(2):381–400, 2018.

[42] European Commission. Special Eurobarometer 487a, 2019. Available at: http://dx.doi.org/10.2838/579882.

[43] D. Hardt. The OAuth 2.0 authorization framework. RFC 6749, RFC Editor, 2012. Available at: http://www.rfc-editor.org/rfc/rfc6749.txt.

[44] S. Landau and T. Moore. Economic tussles in federated identity management. *First Monday*, 17(10), 2012.

[45] Data Portability Cooperation. Telecoms as the "secured data hub" for the digital society, 2019. Available at: https://www.dataportabilitycooperation.org/assets/Telecoms_Secured_Data_Hub.pdf, last accessed: November 27th, 2020.

[46] S. Spiekermann and J. Korunovska. Towards a value theory for personal data. *Journal of Information Technology*, 32(1): 62–84, 2017.

[47] B. Caillaud and B. Jullien. Chicken & Egg: Competition among intermediation service providers. *RAND Journal of Economics*, 34(2):309–328, 2003.

[48] T. Urban, D. Tatang, M. Degeling, T. Holz, and N. Pohlmann. A study on subject data access in online advertising after the GDPR. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, 2019.

A Default request email

We would like to thank Wong and Henderson [11] for providing us with the sample email request from their study. We reused their email to ensure comparability and provide the content below:

From: john@doe.com
Subject: Request for Data Portability under Art. 20 GDPR
Date: Tue, March 10th, 2020 11:34:54
To: Jane Smith <jane.smith@corporate.com>

Dear Sir or Madam,

please supply the personal data about me in a machine−readable format that I am entitled to under Article 20 'Right to data portability' of the General Data Protection Regulation (GDPR) 2016/679.

My personal details related to your organisation are:
Name: John Doe
E−mail address: john@doe.com
Additional information: Username: JohnnyDoe
Should any more information from me be required, please let me know as soon as possible.

It may be helpful for you to know that a request for information under the GDPR should be responded to within one month of receipt. If you do not normally deal with these requests, please pass this email to your Data Protection Officer. If you need advice on dealing with this request, the Office of the German Federal Commissioner for Data Protection and Freedom of Information can assist you and can be contacted on + 49 (0)228−997799−0 or at https://www.bfdi.bund.de/EN/.

Yours faithfully,
John Doe

B Characteristics of exported and imported data

Table 8 gives an overview on what types of data were seen in the data exports and offered for import per industry. Observed data is marked as **bold**, inferred data is marked as *italic*.

Table 8. Main characteristics of exported and imported data (received data, **observed data**, *inferred data*).

Industry	Export	Import
Hardware & Software Manufacturers	Personal Information (name, address, ...) Account Information (incl. billing) Contacts, Calendars, Emails **Account Activity, Usage Data (e.g. login times, clicked links, used devices/browsers, ...)** **Locations** *Advertising interests*	Contacts, Calendars, Emails
Social Communities & Messaging	Personal Information, Account Information, Account Settings Messages, Comments, Bookmarks, Likes, Votes, Stories, ... Upload History, Groups, Contacts **Account Activity, Usage Data** **Searches**, **Locations** *Interests (advertising, jobs, sites)*	Contacts, Calendars Photos Stories Courses
Publishing, Broadcasting & Entertainment	Personal Information, Account Information Playlists **Account Activity, Usage Data** **Playback history**	Videos
Infrastructure Services	Personal Information, Account Information Repositories, Projects **Account Activity** **Locations** *Personalized Ads*	Contacts, Calendars, Emails Projects, Virtual Machines
Leisure, Events & Fitness	Personal Information, Account Information Contacts, Likes, Reviews, Ratings, ... **Account Activity** **Locations** **Fitness & Training Data (e.g. workouts, heart rate, ...)**	Game Data **Locations** **Fitness & Training Data**
Retail, Wholesale & Marketplaces	Personal Information, Account Information, Account Settings Orders, Invoices Comments, Votes, Wish List, ... **Account Activity** **Searches** *Speech Transcripts* *Virtual Assistant Answers*	Products Customer Data
Mobility Services	Personal Information, Account Information, Account Settings Messages, Favorite Places, Routes, ... (as entered by user) Rental information, passenger data **Account Activity** **Searches** **Location** *Speech Transcripts*	Personal Information Calendars
Other Information Services	Personal Information, Account Information Uploaded Files, Bookmarks **Account Activity**	Projects, Designs Customer Data Learning Material
Survey, Marketing & Web Analytics	Personal Information, Account Information Surveys, Mailing Lists **Account Activity**	Contacts, Calendars Surveys, Customer Data **Usage Data (Web Analytics)**
Financial & Legal Services	Personal Information, Account Information Tax Data, Insurance Data (damages, ...), Contracts **Account Activity** *Rating Information (credit rating score)*	Bank Account Data (via Payment Account Switching Service)

C Duration, transmission methods, and authentication types

Figure 4 shows the number of days it took the services to process the data export request per industry.

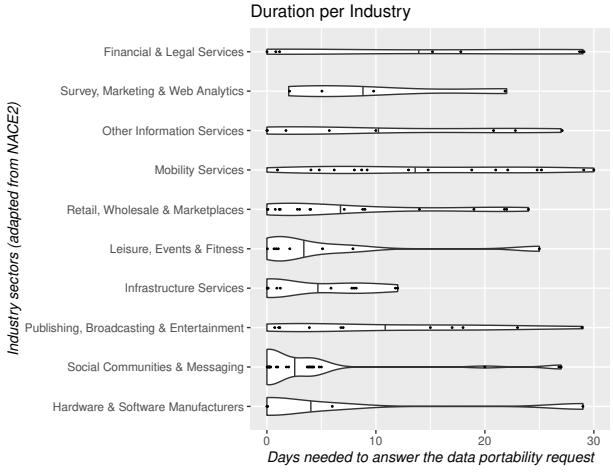

Fig. 4. Duration (in days) per industry sector. Six online services requested the additional 60 days extension and are not illustrated here.

Figure 5 shows the relative frequency of data export transmission methods per industry.

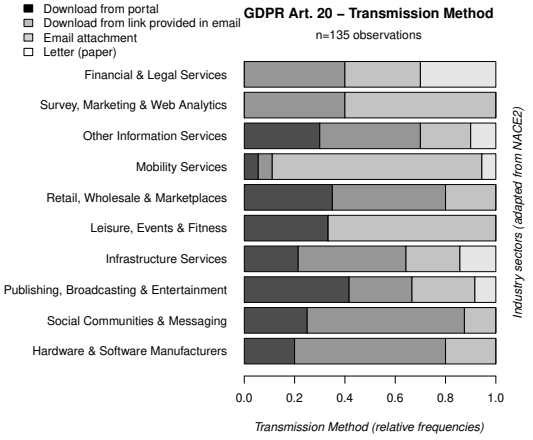

Fig. 5. Transmission methods per industry sector.

Figure 6 shows which types of authentication were used how frequently in each industry.

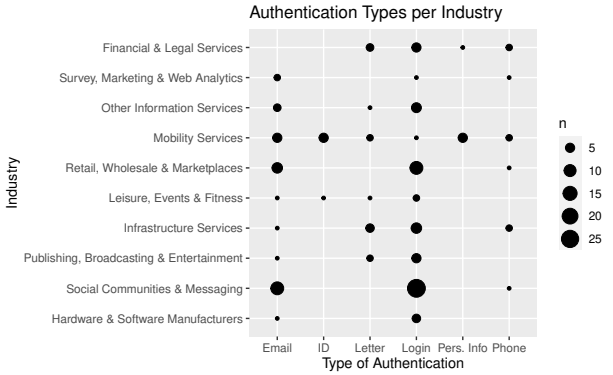

Fig. 6. Authentication types per industry sector.

D Hypothesis 5: Alternative definition of import scope

Merging the import scope classifications *minimal* and *partly* yields the regression results shown in Table 9.

Table 9. Hypothesis 5†: Popularity on Import Scope.

	Dependent variable:
	Import Scope†
log(Alexa)	**−0.224*** **
	(0.076)
Industry	
Social Communities & Messaging	−0.026
	(0.952)
Publishing, Broadcasting & Entertainment	−1.560
	(1.354)
Infrastructure Services	1.359
	(1.008)
Leisure, Events & Fitness	1.757
	(1.158)
Retail, Wholesale & Marketplaces	−1.052
	(1.150)
Mobility Services	−0.373
	(1.206)
Other Information Services	0.262
	(1.058)
Survey, Marketing & Web Analytics	2.415**
	(1.116)
Financial & Legal Services	2.145*
	(1.124)
Observations	190
Akaike Inf. Crit.	241.039

Note: *p<0.1; **p<0.05; ***p<0.01

E Classification of industries by NACE codes

Financial & Legal Services (16 services)

Banks, insurances and payment providers.

6419: Other monetary intermediation
6499: Other financial service activities, except insurance and pension funding n.e.c.
6610: Activities auxiliary to financial services, except insurance and pension funding
6619: Other activities auxiliary to financial services, except insurance and pension funding
6622: Activities of insurance agents and brokers
6910: Legal activities
8291: Activities of collection agencies and credit bureaus

Survey, Marketing & Web Analytics (8 services)

Corporations that provide the service to gather and analyze information on the web or offer survey tools, with which individuals can collect information by themselves.

7311: Advertising agencies

Other Information Services (18 services)

Mix of services in the information industry for which none of the other categories apply, e.g., weather forecasting and translation services.

6200: Computer programming, consultancy and related activities
6201: Computer programming activities
6312: Web portals
7490: Other professional, scientific and technical activities n.e.c.
7820: Temporary employment agency activities
8299: Other business support service activities n.e.c.
9609: Other personal service activities n.e.c.

Mobility Services (26 services)

Airlines, car sharing providers, online travel agencies, and similar services.

5110: Passenger air transport
5510: Hotels and similar accommodation
7710: Renting and leasing of motor vehicles
7711: Renting and leasing of cars and light motor vehicles
7739: Renting and leasing of other machinery, equipment and tangible goods n.e.c.
7911: Travel agency activities
7990: Other reservation service and related activities

Retail, Wholesale & Marketplaces (27 services)

Services that sell tangible products to businesses or consumers and online marketplaces.

1419: Manufacture of other wearing apparel and accessories
4642: Wholesale of clothing and footwear
4647: Wholesale of furniture, carpets and lighting equipment
4649: Wholesale of other household goods
4651: Wholesale of computers, computer peripheral equipment and software
4711: Retail sale in non-specialised stores with food, beverages or tobacco predominating
4719: Other retail sale in non-specialised stores
4741: Retail sale of computers, peripheral units and software in specialised stores
4754: Retail sale of electrical household appliances in specialised stores
4761: Retail sale of books in specialised stores
4771: Retail sale of clothing in specialised stores
4778: Other retail sale of new goods in specialised stores
4791: Retail sale via mail order houses or via Internet

Leisure, Events & Fitness (10 services)

Services aimed at recreational activities, like ticket sellers, fitness and sports apps, and gaming.

7721: Renting and leasing of recreational and sports goods
9200: Gambling and betting activities
9319: Other sports activities
9329: Other amusement and recreation activities

Infrastructure Services (21 services)

Providers and maintainers of traditional and online infrastructure, like hosting, telecommunication, and mail.

3510: Electricity, gas, steam and air conditioning supply
3513: Distribution of electricity
5310: Postal activities under universal service obligation
6120: Wireless telecommunications activities
6190: Other telecommunications activities
6203: Computer facilities management activities
6311: Data processing, hosting and related activities

Publishing, Broadcasting & Entertainment (22 services)

Services where content is passively consumed, like TV stations, streaming providers, and traditional newspapers and magazines.
1811: Printing of newspapers
5811: Book publishing
5813: Publishing of newspapers
5814: Publishing of journals and periodicals
5911: Motion picture, video and television programme production activities
6020: Television programming and broadcasting activities
7722: Renting of video tapes and disks

Social Communities & Messaging (36 services)

Social networks, online forums, career-oriented networks, and messaging apps.
6202: Computer consultancy activities
6209: Other information technology and computer service activities
7312: Media representation

Hardware & Software Manufacturers (6 services)

Developers and publishers of software (including operating systems) and manufacturers of hardware (e.g., mobile devices or computers).
2620: Manufacture of computers and peripheral equipment
5829: Other software publishing