How Portable is Portable? Exercising the GDPR's Right to Data Portability

Janis Wong
School of Computer Science
University of St Andrews
St Andrews, KY16 9SX, UK
jccw@st-andrews.ac.uk

Tristan Henderson
School of Computer Science
University of St Andrews
St Andrews, KY16 9SX, UK
tnhh@st-andrews.ac.uk

UbiComp/ISWC'18 Adjunct, October 8–12, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5966-5/18/10...$15.00
https://doi.org/10.1145/3267305.3274152

Abstract
The new European General Data Protection Regulation has introduced several new rights designed to empower users and regulate imbalances of power between those who collect and control data and those to whom the data refer. In this paper we focus on one particular right, the right to data portability, and examine how it is being implemented. We discuss the responses to 230 real-world data portability requests, and examine the file formats returned and difficulties in making and interpreting requests. We find variation in file formats, not all of which meet the GDPR requirements, and confusion amongst data controllers about the various GDPR rights.

Author Keywords
data protection, GDPR, right to data portability

ACM Classification Keywords
K.4.1 [Public Policy Issues]: Privacy.

Introduction
The introduction of the General Data Protection Regulation (GDPR) [12] has been called "the most significant data reform process in history" [8]. This new law introduces a number of new rights for data subjects, which are intended to rebalance power between citizens and the increasingly sizeable and international companies

that are collecting and exploiting data from them.

As the GDPR has only just come into effect, it is timely to study how this Regulation will work in practice. In this paper we examine one right in particular, the right to data portability (RtDP). We conduct 230 data portability requests, and discuss the success or failure of these requests, the types of data formats used, and the completeness of the requests. We show some of the potential impediments that face data subjects in exercising the RtDP, and discuss future areas for work that could help overcome these problems.

The GDPR

Repealing the Data Protection Directive (DPD) [11], the GDPR came into force on 25th May 2018. The previous DPD was introduced in 1995, and with the rise in international processing of big datasets and increased surveillance both by states and private companies, a new Regulation was required to modernise and harmonise data protection across EU Member States, irrespective of a data subject's nationality or residence.

The GDPR, like the DPD, provides several rights for data subjects (those about whom personal data are collected) to exercise against data controllers (those who collect or determine what these data are used for). While some rights in the GDPR already existed under the DPD, such as the right to access data or the right to rectify data, the Regulation also introduces new rights, such as the focus of this paper, Article 20's RtDP. This right aims to allow data subjects to obtain and reuse their personal data for their own purposes across different services. The RtDP is particularly interesting as it discusses different aspects of technology while attempting to remain technologically neutral. As a whole, the GDPR does not depend on the techniques used (GDPR Recital 15). In the context of data portability, however, certain technologies may be required for its implementation. The RtDP therefore makes for a good case study exploring whether new technology is needed to fully exercise these new powers.

The RtDP requires that "the data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided" (GDPR Article 20(1)) processed based on consent or by contract, and carried out by automated means.

The requirements of EU data protection law are laid out in the GDPR itself, but additional guidance is provided through data protection authorities, either of individual member states such as the UK's Information Commissioner's Office (ICO), or the EU-wide Article 29 Data Protection Working Party (A29WP) or the new European Data Protection Board (EDPB). The A29WP's guidelines on data portability [1] clarify the main elements of data portability, when the RtDP applies, and how portable data must be provided. The A29WP and the GDPR do not prescribe the implementation of how the RtDP should be achieved. The A29WP does, however, describe what data should be included in response to a portability request. The term "provided" in Article 20 is interpreted broadly by the A29WP to include data actively and knowingly provided by the data subject and data gathered by virtue of the use of the data controller's service or servicing device. Notably, this does not include data inferred or derived after analysis. The ICO further provides explanation of Article 20's "structured,

commonly used and machine-readable format" with reference to the Open Data Handbook [20].

When exercising Article 20, data controllers must ensure that the correct data subject is identified. Additional information may be asked to enable the identification of a data subject if there is reasonable doubt about their identity (Article 12(6)). When further information and proof of identity is received, data controllers cannot refuse to act upon the data subject's request (Article 12(2)). Once confirmed, the data controller has up to one month, or up to three months if the size of data requested is significant, and without undue delay to comply (Article 12(3)).

The right to receive portable personal data is not the same as making data interoperable across different platforms. In the EU, interoperability is defined as "the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the exchange of data between their respective ICT systems" (Article 2) [10]. Article 20's requirements for structured and machine-readable formats and clearly-defined metadata are important for data files to be interoperable, but despite pressure from lawyers and academics, mandatory interoperability provisions have not been included in the GDPR [4]. The A29WP's guidance clarifies that interoperability and the production of interoperable systems are only desired outcomes.

Data Portability in Context

Before the GDPR, data portability was grounded in competition law under Article 102 of the Treaty on the Functioning of the European Union for abuse of dominance and exclusionary conduct as well as the Sherman Act and Clayton Act in the US. With the potential for service providers to "lock-in" consumers and make it more difficult for them to leave the platform, data portability is seen as a solution allowing users to move from one service to another. Technical solutions to data portability have been proposed for some time: McCown and Nelson suggested mechanisms for using the Facebook API, browser extension archiving frameworks, and third party web archivers to extract personal data to breakaway from the "walled garden" [18], and Bojars et al. argue that implementing data portability for social networks using Semantic Web technology is technically feasible and comes at almost zero-cost for developers [3]. Beyond advantages for users, Van der Auwermeulen argues that there is also an economic interest for providers to offer data portability [24]. The possibility for portable data encourages users to put more of their personal data onto platforms with the trust that they can transmit it later.

Data portability is not without disadvantages. Swire and Lagos argue that data portability in the then-draft GDPR may reduce consumer welfare as it places excessive burden on small and medium enterprises by disregarding market power [21]. Security challenges may also arise as the complexity of control and process of personal data increases with more portable data [25]. More specifically to data portability as a right, Graef et al. consider how the RtDP clashes with competition law and consumer protection law where data portability is seen as a duty and a form of property-like control respectively [13]. If justifications for data portability are poorly-defined, portability may become a goal in itself with little impact on the protection of personal data [14].

As a new right, data portability has yet to be tried and tested. Ursic suggests that the RtDP could establish

control over personal data transfers, enable (re)use of personal data, enable better understanding of data flows, and allow free development of personality and facilitate equality [23]. To avoid adverse effects on competition and innovation, Engels argues that the nuances of platform market characteristics should be considered during the enforcement and interpretation of Article 20 to prevent barriers to the development of new digital business models [9]. The RtDP should not be seen only from legal perspectives. The IT design community needs to develop technical and organisational safeguards into personal management systems for users, allowing them to better understand how their data is used and maintain agency of their online presence [22]. De Hert et al. also consider the possibilities for building interoperable infrastructures enabling data subjects to bridge the gap between specific services [7].

In exercising data protection rights beyond data portability, a wide body of research was conducted prior to the implementation of the GDPR. Ausloos and Dewitte found that out of 66 data controllers, only 53% of privacy policies were deemed satisfactory and 22% of responses returned were deemed satisfactory [2]. In another study, 106 requests were sent by 7 individuals. It was found that 83% of organisations answered to the access requests, 22% answered subsequent sub-questions, and only 10% specifically identified both the aspect of the data collected and with which organisations data were shared [16]. In the assessment of data controller compliance, using the example of CCTV footage, it was found that data subjects were unable to exercise their rights because responses invoked incorrect or inaccurate legal regulations that restricted data controllers' disclosure obligations. Reliance upon incorrect legislative provisions and delayed responses to deny access were recurring practices across Europe [19].

More recently, stakeholders have begun exploring the universal provision of data portability with interoperability beyond GDPR compliance. The Data Transfer Project is an open-source platform that facilitates direct portability of user data between cloud services by converting proprietary APIs to and from a small number of standardized data formats that can be used by anyone [6]. This partnership between Facebook, Google, Microsoft, and Twitter identifies portability and interoperability as central to innovation, thus promoting user choice, encouraging responsible product development, and maximising benefits for users. Marsden goes further and suggests that interoperability should be applied in law to allow prosumers (users who actively share and produce online content) to move to more prosumer-friendly products if desired [17]. Failure to consider interoperability can also result in dubious competitive advantage, thus generating consumer dissatisfaction, requiring significant legal expenses, and attracting antitrust scrutiny [26].

Data Portability in Practice

Methodology

Our aim was to understand how data controllers have approached Article 20 by studying their response to requests. To exercise the RtDP, a Python program was created to make 230 portability requests. We used e-mail as the transport for the tool; although some controllers offer automated download options, this is far from commonplace (see our results later). On the other hand, e-mail is commonly used by both data subjects and data controllers. A single data subject (the first author) made the requests, and so the data controllers were drawn from a set of organisations who held personal data about the data subject. We categorised these data controllers using the Curlie taxonomy [5]; we omit full results for brevity, but popular categories included "Publications" (12.6%),

"Software" (11.3%), "(Legal) Services" (9.1%), "Clothing" (6.5%), "Non-profit Resources" (5.7%) and "Online Communities" (5.7%).

Data controllers were not told prior to the completion of exercising the RtDP that these requests were made for research purposes so as not to prejudice the responses received. The information required for initiating these requests include the data controller's contact information, expressing the desire to make an RtDP request, and personally identifiable information of the data subject such as name, e-mail, and account usernames. The contact details for Data Protection Officers, or data controllers more generally if a specific data protection related e-mail was not identified, were discovered by manual inspection of websites; typically the privacy policy or terms and conditions pages. Neither the A29WP or ICO guidance provide any example e-mail messages for the RtDP, so a template for the right of access (RoA) was modified to adhere to the requirements under Article 20 [15]. If no response or no indication of acting upon the request was received, a reminder e-mail was sent after three weeks. The study began on the day the GDPR came into effect (25th May 2018) and the data collection process ended on the 26th August 2018.

Results

All data controllers listed a contact e-mail address; 173 contact details were found under the terms and conditions or privacy policy pages, where 104 were specifically privacy- or data-protection-related and 126 were general. Even after looking at these pages, contact information was sometimes only found after being redirected. There was no consistency for finding contact information in the remainder. In 19 cases, the e-mail addresses indicated did not work and an alternative had to be found.

Out of the 230 requests sent, all were successfully delivered apart from two requests; one because the e-mail domain no longer existed and the other because of a bank's specific e-mail security restrictions. A portability request for the latter was submitted again via its mandated web form. Including responses indicating that no personal data were stored, 163 of the requests were successfully completed. Out of the remainder, five asked for the full three months allowed under the GDPR, 29 responded initially but did not react to a follow-up e-mail reminder, and 33 did not respond at all. 50 replied within a week of the request being made, then 22 more within two weeks, and 91 within a month.

In making the requests, 88 data controllers required additional personal data for verifying identity. From these requests, 18 required filling out a designated form, 18 required logging in to personal accounts, 14 required photographic national ID and proof of address, 23 required only an ID, and two required only a proof of address. Questions such as date of birth, most recent bank transaction, and other details related to our interactions within certain services were also asked by 13 data controllers.

A variety of mechanisms were also used by data controllers for sending responses. 126 responses were sent by e-mail of which 19 files were password protected, 18 were retrieved through personal login accounts, 17 were downloaded from an online portal of which eight files were password protected, two were file passwords received by post and 2 were full postal responses. 27 data controllers indicated that they stored no data beyond the e-mail address and correspondence that we provided and so did not have any additional data to provide.

Response file type	No. of responses
Tabular (CSV/XLS/XLSX)	72 (36.5%)
Documentation (HTML/PDF)	35 (17.8%)
Data (JSON/XML)	29 (14.7%)
Text (TEX/TXT)	15 (7.6%)
Word (DOC/DOCX/RTF)	13 (6.6%)
Text in e-mail body	12 (6.1%)
Screenshots (PNG)	6 (3.0%)
Images (JPEG)	4 (2.0%)
Audio (MP4/WAV)	2 (1.0%)
E-mail (EML/MBOX)	2 (1.0%)
Paper scan (PDF)	2 (1.0%)
Calendar (ICS)	1 (0.5%)
Contacts (VCF)	1 (0.5%)
Mapping (KMZ)	1 (0.5%)
Paper (Print-out)	1 (0.5%)
Source code (Repository)	1 (0.5%)

Table 1: File types used in responses to portability requests.

Responses were provided in numerous different types (Table 1), the most popular being tabular CSV or Excel files. Ten data controllers reported that they chose the file formats (CSV, XML, and JSON) suggested by the ICO. As some responses contained more than one file type and more than one file of each type, we normalised the responses to one file of each type per response to calculate the percentages in Table 1.

Despite little mention of security in the RtDP guidance, security had an effect on RtDP responses. Five data controllers required telephone conversations to confirm requester identity. Where password-protected files were used, all data controllers separated the transmission of data files and passwords. In spite of this focus on security, two data controller e-mail responses were lost in transit, suggesting potential vulnerabilities in using e-mail communication for transmitting personal data files. This may be exacerbated by the fact that many file types in Table 1 are often flagged by spam filters.

Within different data controller categories, we saw no pattern in how data were sent. We saw some small tendencies for particular file formats in some categories: Legal Services to Word, Online Communities to Tabular, and Software to Data, but these were not significant.

File Format	Structured?	Commonly Used?	Machine-Readable?
Email body	✗	✓	✗
CSV	✓	✓	✓
DOC/DOCX	✗	✓	✗
EML	✓	✓	✓
HTML	▲	✓	▲
ICS	✓	✓	✓
JPEG	✗	✓	✗
JSON	✓	✓	✓
KMZ	✓	✗	✓
MBOX	✓	✓	✓
MP4	✗	✓	✗
PDF	▲	✓	▲
PNG	✗	✓	✗
RTF	✗	✓	✗
TEX	✓	✓	✓
TXT	▲	✓	▲
VCS	✓	✓	✓
WAV	✗	✓	✗
XLS/XLSX	✓	✓	▲
XML	✓	✓	✓

Table 2: RtDP file format compliance based on "structured, commonly used, and machine-readable" requirements. Most formats either comply (✓) or do not comply (✗) with the ICO definitions but some are ambiguous (▲).

There was little consensus on what is required for full compliance with Article 20. Importantly, although all data controllers indicated that their responses comply with Article 20, it is questionable whether data were always delivered in a "structured, commonly used, machine-readable format". Table 2 lists file-format compliance based on the ICO's "structured, commonly used, and machine-readable format" definitions. We consider a format to be "structured" where structural relation between elements are explicit, "commonly used" where formats are identifiable beyond the data controller's usage, and "machine-readable" where data can be identified, recognised, and extracted in processing. It is not always clear whether a format meets all the criteria. For example, CSV files are compliant because there are structural relationships between elements within tabular data, is a commonly used format, and can be processed by computers. By contrast, although HTML is a commonly used format for web pages, it is only structured and machine-readable with markup. Formats may fall into grey areas (marked ▲) where specific technical processing or levels of metadata beyond its default format standard are required. For instance PDF files may be machine-readable if they contain text, but not so much if they contain images or scans. 40% of the file formats were identified as fully-compliant: CSV, EML, ICS, JSON, MBOX, TEX, VCS, and XML. Even if the personal data held by data controllers were structured and machine-readable during processing, the extraction process made some files non-compliant. For instance, with screenshots and paper scans, although the information was provided, the choice of format made machine-readability difficult.

Finally, we observed some confusion between the different rights in the GDPR. Four data controllers misunderstood our RtDP request as exercising the RoA. Two went further to suggest that we make a RoA request instead so that we can have more personal data. Four RtDP requests were conflated with the right to erasure and the right to restrict processing. Four data controllers explicitly noted that our RtDP requests were the first they had received. After completing the requests, one data controller asked us for feedback on what we thought about the process and compliance. One data controller initially said that no data were stored but came back a month later with personal data that were previously not revealed. Two data controllers mentioned that they were unsure whether certain data were required under the GDPR. One data controller claimed that their system could not provide information in a machine-readable format. Our study also caused one data breach, where a response to our RtDP included personal data of other data subjects.

Making Portable Data Portable

Our results show problems around portability both for data controllers, who may misunderstand requirements or provide data in inappropriate or incomplete formats, and data subjects, who may be unable to verify their identity or verify the veracity of the data returned by a controller. Our study is still at an early stage and raises several possibilities for future work.

We have only examined file formats so far, but this has shown that it is difficult to determine what files are compliant and what files are not. One avenue for future work is to examine the content of our responses to further clarify how and when certain formats comply, what content should be included in RtDP responses, and what may be necessary to ensure compliance.
Technically-advanced definitions or standards for "structured, commonly used, and machine-readable" could be provided so that it becomes actionable, allowing

appropriate tests to be designed.

We have examined data transfers from data controller to data subject, but Article 20 also offers a mechanism for data subjects to request that their data be transferred directly to another data controller. Future work is needed to explore how this can enable data subjects to use Article 20 as a mechanism for data protection, by making RtDP requests to transfer interoperable data to other controllers. But given the state of the art shown in our current work, we suggest that such a future study wait until data controllers have become more familiar with Article 20.

Building upon the metadata from the RtDP responses received, more empirical work can be done to assess the feasibility of interoperability. As interoperability is not required by the GDPR, there is no obligation to apply transmitted data from one service to another. Portable data itself can be transmitted but the spirit and value of portability is lost if data is not meaningfully reused by other data controllers. Additionally, existing methods for verifying data subject identities such as phone calls and the necessity to clarify RtDP data required may act as hurdles for enabling interoperable data. Based on the metadata received from RtDP responses, the interoperability of specific categories, such as social networks, can be explored. Challenges for interoperability, such as the technological infrastructures required, the problems with existing verification processes, and what categories of data controllers can be made interoperable, should be examined to ensure that portable data is legally and technologically portable.

Finally, we are also exploring technological routes for helping data subjects to exercise data protection rights such as the RtDP, integrating identification and secure storage to make it easier to make and verify requests.

Conclusion

In this paper we examined the GDPR's RtDP by making 230 real-world requests. We found a variety of file formats being returned by data controllers, some of which may not comply with the obligations in Article 20, and some confusion between the various rights in the GDPR on the part of data controllers. Future work is needed to help both data controllers and data subjects understand and exercise these rights.

References

[1] Article 29 Data Protection Working Party. 2017. 'Guidelines on the right to data portability'. WP 242 rev.01. (5th Apr. 2017).

[2] J. Ausloos and P. Dewitte. 2018. 'Shattering one-way mirrors — data subject access rights in practice'. *International Data Privacy Law*, 8, 1, (Feb. 2018), 4–28. DOI: 10.1093/idpl/ipy001.

[3] U. Bojars, A. Passant, J. Breslin and S. Decker. 2008. 'Social networks and data portability using semantic web technologies'. In *2nd Workshop on Social Aspects of the Web*.

[4] I. Brown and C. T. Marsden. 2013. *Regulating Code*. MIT Press, Cambridge, MA, USA.

[5] The Curlie Directory. 2018. Retrieved 25/08/2018 from http://curlie.org/.

[6] The Data Transfer Project. 2018. Retrieved 28/07/2018 from https://datatransferproject.dev/.

[7] P. De Hert, V. Papakonstantinou, G. Malgieri, L. Beslay and I. Sanchez. 2018. 'The right to data portability in the GDPR: towards user-centric interoperability of digital services'. *Computer Law & Security Review*, 34, 2, (Apr. 2018), 193–203. DOI: 10.1016/j.clsr.2017.10.003.

[8] L. Edwards. 2018. 'Data protection: Enter the General Data Protection Regulation'. In *Law, Policy and the Internet*. L. Edwards, editor. Hart Publishing, London. DOI: 10.2139/ssrn.3182454.

[9] B. Engels. 2016. 'Data portability among online platforms'. *Internet Policy Review*, 5, 2, (June 2016). DOI: 10.14763/2016.2.408.

[10] European Union. 2015. 'Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens (ISA2 programme) as a means for modernising the public sector'. *Official Journal of the European Union*, L318, (4th Nov. 2015), 1–16.

[11] European Union. 1995. 'Directive of the European Parliament and the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data'. *Official Journal of the European Union*, L281, (23rd Nov. 1995), 1–20.

[12] European Union. 2016. '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)'. *Official Journal of the European Union*, L119, (4th May 2016), 1–88.

[13] I. Graef, D. Clifford and P. Valcke. 2018. 'Fairness and enforcement: bridging competition, data protection and consumer law'. *International Data Privacy Law (forthcoming)*, (July 2018).

[14] I. Graef, M. Husovec and N. Purtova. 2017. 'Data portability and data control: lessons for an emerging concept in EU law'. *International Data Privacy Law (forthcoming)*, (Nov. 2017). DOI: 10.2139/ssrn.3071875.

[15] Information Commissioner's Office. 2018. *Your right of access*. Retrieved 01/07/2018 from https://ico.org.uk/your-data-matters/your-right-of-access/.

[16] R. Mahieu, H. Asghari and M. van Eeten. 2017. 'Collectively exercising the right of access: individual effort, societal effect'. In *GigaNet (Global Internet Governance Academic Network) Annual Symposium*. (Dec. 2017). DOI: 10.2139/ssrn.3107292.

[17] C. Marsden. 2018. 'Prosumer law and network platform regulation: the long view towards creating offdata'. *Georgetown Law Technology Review*, 2, 2, 376–398.

[18] F. Mccown and M. L. Nelson. 2009. 'What happens when facebook is gone?' In *9th ACM/IEEE-CS Joint Conference on Digital libraries*. (June 2009), 251–254. DOI: 10.1145/1555400.1555440.

[19] C. Norris, P. de Hert, X. L'Hoiry and A. Galetta, editors. 2017. *The Unaccountable State of Surveillance*. Springer, Cham. DOI: 10.1007/978-3-319-47573-8.

[20] Open Knowledge International. 2015. *The open data handbook*. Retrieved 25/07/2018 from http://www.opendatahandbook.org/guide/en/.

[21] P. Swire and Y. Lagos. 2013. 'Why the right to data portability likely reduces consumer welfare: antitrust and privacy critique'. *Maryland Law Review*, 72, 2, 335–380. DOI: 10.2139/ssrn.2159157.

[22] L. Urquhart, N. Sailaja and D. McAuley. 2018. 'Realising the right to data portability for the domestic Internet of things'. *Personal and Ubiquitous Computing*, 22, 2, (Apr. 2018), 317–332. DOI: 10.1007/s00779-017-1069-2.

[23] H. Ursic. 2018. 'Unfolding the new-born right to data portability: four gateways to data subject control'. *SCRIPT-ed*, 15, 1, (Aug. 2018), 42–69. DOI: 10.2966/scrip.150118.42.

[24] B. Van der Auwermeulen. 2017. 'How to attribute the right to data portability in Europe: a comparative analysis of legislations'. *Computer Law & Security Review*, 33, 1, (Feb. 2017), 57–72. DOI: 10.1016/j.clsr.2016.11.012.

[25] S. Weiss. 2009. 'Privacy threat model for data portability in social network applications'. *International Journal of Information Management*, 29, 4, 249–254. DOI: 10.1016/j.ijinfomgt.2009.03.007.

[26] N. Zingales. 2015. 'Of coffee pods, videogames, and missed interoperability: reflections for EU governance of the Internet of Things'. (2015). DOI: 10.2139/ssrn.2707570.