THE RIGHT TO DATA PORTABILITY IN PRACTICE: EXPLORING THE IMPLICATIONS OF THE TECHNOLOGICALLY NEUTRAL GDPR

Janis Wong, Tristan Henderson

School of Computer Science, University of St Andrews, UK

{jccw,tnhh}@st-andrews.ac.uk

22 May 2019

This article has been accepted for publication in International Data Privacy Law Published by Oxford University Press.

KEY POINTS

- The European General Data Protection Regulation (GDPR) introduces one new data subject right, Article 20's right to data portability (RtDP). The RtDP aims to allow data subjects to obtain and reuse their personal data for their own purposes across different services.

- We investigate the RtDP by making 230 real-world data portability requests across a wide range of data controllers. The RtDP is interesting to study as it operates under a framework that aims to be technologically neutral while requiring specific technologies for implementation. Our objective is to assess the ease of the RtDP process from the perspective of the data subject and to examine the file formats returned by data controllers.

- From our results, including responses indicating that no personal data were stored, only 172 (74.8%) of RtDP requests were successfully completed. However, compliance with the GDPR varied where not all file formats meet the GDPR requirements. There was also confusion amongst data controllers about data subject rights more generally.

- Based on our observations, we revisit the current guidance for data portability. We suggest new technical definitions to clarify how data should be made portable and determine the appropriateness of certain file formats for different data types.

- We suggest recommendations and future work for various stakeholders to address the legal implications derived from our study. This includes discussing possibilities for new data portability standards and codes, conducting further empirical research, and building technological solutions to ensure that the RtDP can be better understood in theory and exercised in practice.

KEYWORDS

Data portability; data subject rights; General Data Protection Regulation; technological neutrality

I. INTRODUCTION

The introduction of the General Data Protection Regulation (GDPR)[1] has been called 'the most significant data privacy reform process in history'.[2] This new law reinforces existing data subject rights in an attempt to rebalance power between citizens and the increasingly sizeable and international companies that are collecting and exploiting data from them.

As the GDPR only came into effect recently on the 25 May 2018, it is timely to study how this Regulation works in practice. In this paper, we examine the new data subject right introduced under the GDPR, Article 20's the right to data portability (RtDP). Described by the European Data Protection Supervisor (EDPS) as 'the gateway in the digital environment to the user control which individuals are now realising they lack',[3] the RtDP provides the right for data subjects to receive personal data concerning him or her and the right to transmit those data from one data controller to another. The RtDP is interesting to study as it operates under a framework that aims to be technologically neutral to maintain reasonable longevity of the law. In Section II, we explore the historical background, current developments, and existing research on how data portability sits within the GDPR. Traditionally grounded in competition and consumer law, the RtDP is also the first of its kind to be included in data protection law. As a new right, it is yet to be seen whether data portability can be used as a means for protecting the processing of data subjects' personal data. However, given that data portability requires specific technologies for implementation, little

[1]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] OJ L119/1.
[2]Lilian Edwards, 'Data Protection: Enter the General Data Protection Regulation' in Lilian Edwards (ed), *Law, Policy and the Internet* (Hart Publishing 2018) 77.
[3]European Data Protection Supervisor, 'EDPS recommendations on the EU's options for data protection reform' (2015) C301 OJ 1.

guidance has been provided to data controllers with regards to compliance. To see how the RtDP works in practice, we then conduct 230 data portability requests to a broad range of data controllers and discuss the successes and failures of making these requests as data subjects. With the responses received, we assess the types of data formats used, and the completeness and appropriateness of the RtDP requests. We show some of the potential impediments that data subjects and data controllers may face in exercising and complying with the RtDP respectively. Then, we revisit the definitions of data portability terminology based on the legal implications identified from the GDPR's technological neutral framework in application to the RtDP. Finally, we discuss future areas for work in Section V, where different stakeholders could help to further clarify the RtDP and suggest ways to overcome these obstacles.

II. GDPR ARTICLE 20 – THE RIGHT TO DATA PORTABILITY

Replacing the Data Protection Directive (DPD),[4] the GDPR came into force on the 25 May 2018. The 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.[5]

The GDPR, like the DPD, provides several rights with regards to personal data for data subjects as 'identified or identifiable natural'[6] persons to exercise against data controllers as those who determine 'the purposes and means of the processing of personal data'.[7] The Regulation represents a significant change to existing data protection law. It strengthens these existing data subject rights

[4]Directive 95/46/EC of the European Parliament and of 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 [1995] OJ L281/31.
[5]Reg 2016/679 (n 1) rec 14.
[6]Reg 2016/679 (n 1) art 4(1).
[7]Reg 2016/679 (n 1) art 4(7).

by clarifying data controller responsibilities and explaining the rights offered to data subjects. Unlike the DPD, data controllers' responses to data subject rights requests shall be provided free of charge[8] unless manifestly unfounded or excessive.[9] When exercising the RtDP, data controllers must take appropriate measures to ensure that data subject rights are correctly exercised under Article 12. Additional information about the data subjects may be asked to enable their identification if there is reasonable doubt about their identity.[10] When further information and proof of identity is received, data controllers cannot refuse to act upon the data subject's request.[11] Once confirmed, the data controller has up to one month, or up to three months if the complexity and number of requests are significant, to provide information and must do so without undue delay.[12]

While some rights in the GDPR already existed under the DPD, such as the right to access and the right to rectify data, the Regulation also introduces one new right and 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.

Article 20, with the parts important for our understanding of the practical application of the RtDP highlighted, states:

1. 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, where:

[8]Reg 2016/679 (n 1) art 12(5).
[9]Reg 2016/679 (n 1) art 12(2).
[10]Reg 2016/679 (n 1) art 12(6).
[11]Reg 2016/679 (n 1) art 12(2).
[12]Reg 2016/679 (n 1) art 12(3).

(a) the processing is based on **consent** pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a **contract** pursuant to point (b) of Article 6(1); and

(b) the processing is carried out by **automated means**.

2. In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the **right to have the personal data transmitted directly from one controller to another, where technically feasible**.

3. The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.

4. The right referred to in paragraph 1 shall not adversely affect the rights and freedoms of others.

The RtDP offers data subjects 'the right to receive the personal data concerning him or her' and 'the right to have the personal data transmitted directly from one controller to another, where technically feasible' under GDPR Article 20(1) and Article 20(2) respectively. Under the RtDP, data subjects can only receive personal data 'which he or she has provided to a controller'[13] where the processing is based on consent or by contract[14] and is carried out by automated means.[15]

The RtDP is particularly interesting to study as it relates to different aspects of technology while attempting to remain technologically neutral.[16] As a whole, the GDPR does not depend on the

[13]Reg 2016/679 (n 1) art 20(1).
[14]Reg 2016/679 (n 1) art 20(1)(a).
[15]Reg 2016/679 (n 1) art 20(1)(b).
[16]Reg 2016/679 (n 1) rec 15.

techniques used.[17] In the context of data portability, however, certain technologies, such as for processing and extracting data, converting such data to specific file formats, and transporting RtDP responses back to data subjects, may be required for its implementation. Given both the introduction of the GDPR and, as a result, Article 20, the RtDP therefore makes for a timely case study for exploring whether new technology is needed to fully exercise these new powers.

Implementing the RtDP

The requirements of EU data protection law are laid out in the GDPR itself, but additional guidance is provided through Data Protection Authorities (DPAs), either by the EU-wide Article 29 Data Protection Working Party (A29WP), now replaced by the new European Data Protection Board (EDPB), or by individual Member States through regulators such as the UK's Information Commissioner's Office (ICO). While only the Regulation alone is a binding legislative act, the guidance produced by DPAs plays a key role. Any guidance produced by authorities can be referred to during enforcement.[18]

The A29WP did not and the GDPR does not prescribe how the implementation of the RtDP should be achieved. The A29WP did, however, describe which data should be included in response to a portability request. The A29WP's guidelines on data portability clarified the main elements of data portability, when the RtDP applies, and how portable data must be provided.[19] This was endorsed by the EDPB when the GDPR came into force on the 25 May 2018.[20] The term 'provided' in Article 20 was interpreted broadly by the A29WP to include data actively and knowingly

[17]Reg 2016/679 (n 1) rec 15.
[18]Commission, 'Communication from the Commission to the European Parliament and the Council: Stronger protection, new opportunities - Commission guidance on the direct application of the General Data Protection Regulation as of 25 May 2018' COM(2018) 43.
[19]Article 29 Data Protection Working Party, 'Guidelines on the right to data portability' (WP 242 rev.01, 2017).
[20] European Data Protection Board, 'Endorsement of GDPR WP29 guidelines by the EDPB' (Endorsement 1/2018).

provided by the data subject as well as data gathered by virtue of the use of the data controller's service or servicing device.[21] Notably, this does not include data inferred or derived after analysis.[22] In the UK, the ICO provides guidance on what kinds of data the RtDP relates to, how data subjects should ask data controllers for their portable data, when a portability request should be made, and how concerns about the methods in which data controllers have handled personal data can be made.[23] For data controllers and organisations, the ICO also clarifies their responsibilities and limits with regards to secure transmission, the transmission of personal data to another controller, how to respond to data portability requests, and how to comply with them.[24]

Given that the GDPR itself does not define Article 20's 'structured, commonly used and machine-readable format',[25] the A29WP clarified that the terms 'are a set of minimal requirements that should facilitate the interoperability of the data format provided by the data controller.'[26] Only 'machine-readable' is formally defined by the EU as a 'file format structured so that software applications can easily identify, recognize and extract specific data, including individual statements of fact, and their internal structure.'[27] Within the UK's data protection enforcement regime, the ICO explains these terms with reference to the Open Data Handbook. 'Structured data' is 'data where the structural relation between elements is explicit in the way the data is stored on a computer disk.'[28] Thus, a structured format is one from which software can extract specific and known

[21]Article 29 Data Protection Working Party (n 19) 8.
[22]Article 29 Data Protection Working Party (n 19) 9.
[23]Information Commissioner's Office, 'Your right to data portability' (9 October 2018) https://ico.org.uk/your-data-matters/your-right-to-data-portability/ accessed 25 January 2019.
[24]Information Commissioner's Office, 'Right to data portability for organisations' (4 January 2019) https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/ accessed 25 January 2019.
[25]Reg 2016/679 (n 1) art 20(1).
[26]Article 29 Data Protection Working Party (n 19) 13.
[27]Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information Text with EEA relevance [2013] OJ L175/1, rec 21.
[28]Open Data Handbook (2018) https://opendatahandbook.org/glossary/en/terms/structured-data/ accessed 25 January 2019.

elements of data. 'Commonly used' is described as 'widely-used and well-established'.[29] 'Machine-readable' is 'data in a data format that can be automatically read and processed by a computer, such as Comma Separated Values (CSV), JavaScript Object Notation (JSON), Extensible Markup Language (XML), etc. Machine-readable data must be structured data.'[30] While there is no legal requirement for specific formats, the ICO suggests that CSV, XML, and JSON files are acceptable.[31] A CSV file is a 'standard format for spreadsheet data. Data is represented in a plain text file, with each data row on a new line and commas separating the values on each row. As a very simple open format it is easy to consume and is widely used for publishing open data.'[32] This differs from the definition provided by the Open Data Handbook, which focuses on defining the format as one where the structure of the file must be respected and documentation of individual fields are accurate to ensure that data in CSV files are useful.[33] An XML file is defined as a 'simple and powerful standard for representing structured data'[34] by the ICO and is a file format that 'allows developers to write parts of the documentation in with the data without interfering with the reading of them.'[35] A JSON file is a:

> simple but powerful format for data. It can describe complex data structures, is highly machine-readable as well as reasonably human-readable, and is independent of platform and programming language, and is therefore a popular format for data interchange between programs and systems.[36]

---

[29]Information Commissioner's Office, 'Your right to data portability' (n 22).
[30]Open Data Handbook (2018) https://opendatahandbook.org/glossary/en/terms/machine-readable/ accessed 25 January 2019.
[31]Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR)' (22 January 2019) https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/ accessed 25 January 2019.
[32]Information Commissioner's Office, 'Your right to data portability' (n 22).
[33]Open Data Handbook (2018) http://opendatahandbook.org/guide/en/appendices/file-formats//#comma-separated-files accessed 25 January 2019.
[34]Information Commissioner's Office, 'Your right to data portability' (n 22).
[35]Open Data Handbook (2018) http://opendatahandbook.org/guide/en/appendices/file-formats//#xml accessed 25 January 2019.
[36]Information Commissioner's Office, 'Right to data portability for organisations' (n 23).

The Open Data Handbook describes the file format as 'generally easier for computers to process than others.'[37] While the definitions are based on those from open data, RtDP file formats can be open or proprietary and do not necessarily have formal standards. The ICO links to further W3C resources for understanding XML[38] and JSON[39] files.

Regarding interoperability, Article 20's requirements for structured and machine-readable formats and clearly-defined metadata are important. Despite pressure from lawyers and academics,[40] mandatory interoperability provisions have not been included in the GDPR.[41] The A29WP's guidance clarified that interoperability and the production of interoperable systems are only desired outcomes.[42] The right to receive portable personal data is not the same as making data interoperable across different platforms. While interoperability is not explained in the GDPR, following the EU Decision 2015/2240, 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.[43]

---

[37]Open Data Handbook (2018) http://opendatahandbook.org/guide/en/appendices/file-formats//#json accessed 25 January 2019.
[38]W3C (26 November 2008) http://www.w3.org/TR/2008/REC-xml-20081126/ accessed 25 January 2019.
[39]W3C (March 2014) https://tools.ietf.org/html/rfc7159 accessed 25 January 2019.
[40]Ian Brown and Christopher T Marsden, *Regulating Code* (MIT Press 2013) 42.
[41]Inge Graef, 'Mandating portability and interoperability in online social networks: Regulatory and competition law issues in the European Union' (2015) 39(6) Telecommunications Policy 502 (Special Issue on ITS 2013 Florence) DOI: 10.1016/j.telpol.2015.04.001.
[42]Article 29 Data Protection Working Party (n 19) V.
[43]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 [2015] OJ L318/1, Article 2.

Data Portability in Context

Although data portability is a new GDPR right, the concept itself was explored as early as 2007. Data portability gained traction through the Data Portability Project, an advocacy and evangelism organisation dedicated to 'promoting the adoption and application of data portability' with the belief that data portability enables a borderless experience for people to move easily between network services, reusing data while controlling user privacy.[44] The group became inactive in 2016 but the concept has since developed into legal discourse.

Before the GDPR, data portability was grounded in competition law under Article 102 of the Treaty on the Functioning of the European Union (TFEU) for abuse of dominance and exclusionary conduct[45] as well as the Sherman Act[46] and Clayton Act[47] 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.[48]

The introduction of data portability in data protection was seen as a way of modernising the law to better protect data subjects' personal data in our new digital realities. Data portability, argues Zanfir, encourages the free development of human personality, where the means to achieve this goal are technical processes directly linked to the protection of informational privacy and assuring fair competition.[49] Lynskey emphasises that the RtDP has normative values in ensuring individual control over personal data.[50] Ursic goes further to suggest that the RtDP could establish control

---

[44]Data Portability Project (2016) https://www.dataportability.org accessed 25 January 2019.
[45]Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C326/47 (TFEU).
[46]'An act to protect trade and commerce against unlawful restraints and monopolies (Sherman Antitrust Act of 1890) 26 Stat 209' [1890] 51st United States Congress.
[47]'Clayton Antitrust Act of 1914 38 Stat 730' [1914] 63rd United States Congress.
[48]European Commission, 'Privacy Platform event: Competition and Privacy in Markets of Data' (Joaquń Almunia,26 November 2012) http://europa.eu/rapid/press-releaseSPEECH-12-860en.htm accessed 25 January 2019.
[49]Gabriela Zanfir, 'The right to Data portability in the context of the EU data protection reform' (2012) 2(3) International Data Privacy Law 149 DOI: 10.1093/idpl/ips009.
[50]Orla Lynskey, 'Aligning data protection rights with competition law remedies? The GDPR right to data portability.' [2017] European Law Journal.

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.[51] 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.[52] 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.[53]

Data portability has benefits beyond data protection alone. McCown and Nelson suggested mechanisms for using the Facebook API, browser extension archiving frameworks, and third party web archivers to extract personal data to break away from the 'walled garden'.[54] 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.[55] Beyond advantages for data subjects, Van der Auwermeulen argues that there is also an economic interest for providers to offer data portability.[56] The possibility for portable data encourages data subjects as users to put more of their personal data onto platforms with the trust that they can transmit them later. With Information Technology design considerations, technical and organisation safeguards in personal management

---

[51]Helena Ursic, 'Unfolding the New-Born Right to Data Portability: Four Gateways to Data Subject Control' (2018) 15(1) SCRIPT-ed 42 DOI: 10.2966/scrip.150118.42.

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

[53]Inge Graef, Damian Clifford, and Peggy Valcke, 'Fairness and enforcement: bridging competition, data protection, and consumer law' (2018) 8(3) International Data Privacy Law 200 DOI: 10.1093/idpl/ipy013.

[54]Frank McCown and Michael L Nelson, 'What happens when facebook is gone?' (JCDL '09: 9th ACM/IEEE-CS joint conference on Digital libraries, Austin, 2009) DOI: 10.1145/1555400.1555440.

[55]Uldis Bojars and others, 'Social Networks and Data Portability using Semantic Web technologies' (2nd Workshop on Social Aspects of the Web,2008).

[56]Barbara Van der Auwermeulen, 'How to attribute the right to data portability in Europe: A comparative analysis of legislations' (2017) 33(1) Computer Law & Security Review 57 DOI: 10.1016/j.clsr.2016.11.012.

systems can be developed, allowing users to better understand how their data is used and maintain agency of their online presence.[57]

As international data and information transfers become the norm with frameworks such as the EU–US Privacy Shield,[58] data portability has become a global necessity where personal data is often collected, processed, and stored across borders. Making data portable has the potential to significantly empower individuals living in data-driven societies [59] that prioritise collecting and exploiting personal data at the expense of our data protection, privacy, and fundamental freedoms. As a result, while this paper focuses on the RtDP in the UK context given the geographical location of the authors, other jurisdictions and jurisprudence are considered. With data portability increasingly seen as a mechanism for ensuring greater control of data subjects' personal data, more jurisdictions are considering it in their jurisprudence. In Australia, the proposed Consumer Data Right, put forward by the Australian Productivity Commission, is broader than the RtDP as it covers data that is not strictly personal and extends beyond data provided by the data subject, regardless of the basis for the initial data collection.[60] The new California Consumer Privacy Act (CCPA) of 2018 also introduces data portability to the USA. Although the CCPA is similar to GDPR's Article 20, its remit is broader with less precise technical requirements. The CCPA states that where a:

> business that receives a verifiable consumer request from a consumer to access personal information shall promptly take steps to disclose and deliver, free of charge to the consumer, the personal information required … may be delivered by mail or electronically, and if provided

[57]Lachlan Urquhart, Neelima Sailaja, and Derek McAuley, 'Realising the right to data portability for the domestic Internet of things' (2018) 22(2) Personal and Ubiquitous Computing 317 DOI: 10.1007/s00779-017-1069-2.
[58]Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-US Privacy Shield [2016] OJ L4176/1.
[59]Alex Pentland, 'The Data-Driven Society' (2015) 309 Scientific American 78.
[60]Samson Yoseph Esayas and Angela Daly, 'The Proposed Australian Consumer Data Right: A European Comparison' (2018) 3(2) European Competition and Regulatory Law Review 57.

electronically, the information shall be in a portable and, to the extent technically feasible, in a readily useable format that allows the consumer to transmit this information to another entity without hindrance.[61]

Exercising Data Subject Rights

As a new right, data portability has yet to be tried and tested empirically. In order to understand data portability within the GDPR, data protection needs to be contextualised under the repealed legislation. Prior to the GDPR, the DPD's Right of Access (RoA)[62] was the primary way that data subjects could exercise their rights for protecting the processing of their personal data. While it was a noble goal, compliance and enforcement have been found to be weak.

In exercising the RoA, different Member States had different ways of implementing the right. Norris et al. conducted extensive research on CCTV footage to assess the similarities and differences of exercising the RoA across Europe.[63] First assessing compliance, 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.[64] Regarding enforcement, RoA was further complicated by state legislation. As the previous data protection legislation was a directive and not automatically applied into national law, exercising the same right for the same situation could differ between EU countries. This placed significant burden on Member States and DPAs to establish their own enforcement procedures. 36% of data

[61]'The California Consumer Privacy Act of 2018 Assembly Bill 375' [2018] California Legislative Information.
[62]Dir 95/46/EC (n 4) art 12.
[63]Clive Norris and others (eds), *The Unaccountable State of Surveillance* (vol 34, Springer International Publishing 2017) DOI: 10.1007/978-3-319-47573-8.
[64]Norris and others (n 61) 429.

protection complaints remained unresolved, ranging from 100% resolution in the UK to none in Austria, Hungary, and Slovakia.[65] The cost for filing complaints also differed.[66]

Studies on compliance within data protection frameworks found that data controllers had different attitudes towards what was sufficient. Conducting empirical research a year before the GDPR's implementation, Ausloos and Dewitte found that out of 66 data controllers, only 53% of privacy policies and 22% of responses returned were deemed satisfactory.[67] In another study by Mahieu et al., 106 requests were sent by 7 individuals. 83% of organisations answered to RoA requests, 22% answered subsequent sub-questions, and only 10% identified both the aspect of data collection and which organisations personal data were shared with.[68] Broader research conducted by Kamarinou et al. into the treatment of terms and privacy policies by cloud service providers revealed inconsistencies in detail, lack of transparency about third party storage and processing of personal data, and difficulties for individuals to keep track of changes in the providers' terms and privacy policies.[69] A study of online tracking companies after the GDPR came into force found that in exercising the RoA for 36 companies, only 32 companies (89%) replied within the period defined by law and only 21 (58%) finished the process by the deadline set in the GDPR.[70] Although the authors do not exercise the RtDP, they assess the responses to their RoA requests through the lens of data portability by incorporating it into their analysis without making separate requests. Urban et al. found that most responses were either machine-readable or human-readable. Only one

[65]Norris and others (n 61) 452.
[66]Norris and others (n 61) 453.
[67]Jef Ausloos and Pierre Dewitte, 'Shattering one-way mirrors — data subject access rights in practice' (2018) 8(1) International Data Privacy Law 4 DOI: 10.1093/idpl/ipy001.
[68]Rene Mahieu, Hadi Asghari, and Michel van Eeten, 'Collectively Exercising the Right of Access: Individual Effort, Societal Effect' (GigaNet,2017) DOI: 10.2139/ssrn.3107292.
[69]Dimitra Kamarinou, Christopher Millard, and W Kuan Hon, 'Cloud privacy: an empirical study of 20 cloud providers' terms and privacy policies–Part II' (2016) 6(3) International Data Privacy Law 170 DOI: 10.1093/idpl/ipw004.
[70]Tobias Urban and others, 'The Unwanted Sharing Economy: An Analysis of Cookie Syncing and User Transparency under GDPR' [2018] arXiv:1811.08660.

response they received back were both. A more recent study by Talend found that 72 (70%) companies out of 103 could not fulfil RoA and RtDP requests within the one month time limit.[71]

Although no empirical research has been specifically done for the RtDP, independent third party tools have been developed that may be useful for exercising data subject rights. When created following GDPR requirements, tools can provide guidance for the rights of data subjects without reading the law. As the tools are developed by third parties, data subjects do not need to self-manage the process. For data controllers, tools do not absolve them of their responsibility to inform data subjects about their rights but can lower the resources dedicated to replying to requests. There may be greater certainty and consistency with the standardisation of responses. Data controllers interact with fewer individuals since all requests will be semi-managed by the tool, bridging the communication gap between data subjects and data controllers. For the RtDP, some tools include Datastreams[72] (a consent manager for data controllers and a data stream manager for data processors), the Data Transfer Project[73] (an open-source, service-to-service platform that facilitates direct portability of user data between cloud services by converting proprietary APIs to and from a small set of standardised data formats, founded by Facebook, Google, Microsoft, and Twitter), Fair&Smart[74] (an application that helps French data subjects claim GDPR rights, regain control of their privacy, and decide who to trust with their personal data, while supporting GDPR compliance and management services for data controllers), My Data Done Right[75] (a project that helps Dutch data subjects exercise the RoA and RtDP), OpenGDPR[76] (an open-source common framework that

[71]Talend (13 September 2018) https://www.talend.com/about-us/press-releases/the-majority-of-businesses-are-failing-to-comply-with-gdpr-according-to-new-talend-research/ accessed 25 January 2019.
[72]Datastreams (2019) https://www.datastreams.io/ accessed 25 January 2019.
[73]The Data Transfer Project, 'Data Transfer Project' (28 July 2018) https://datatransferproject.dev/ accessed 24 August 2018.
[74]Fair & Smart (2019) https://www.fairandsmart.com/ accessed 25 January 2019.
[75]My Data Done Right (2019) https://mydatadoneright.eu/ accessed 25 January 2019.
[76]OpenGDPR (12 June 2018) https://github.com/opengdpr/opengdpr accessed 25 January 2019.

has a machine-readable specification, allowing data controllers and data processors to communicate and manage data subject requests in a uniform, scalable, and secure manner), and Port.im[77] (an application that provides data controllers with the possibility of connecting applications together, creating GDPR compliant agreements, communicating how personal data is stored to data subjects, and taking control of the data stored). While tools help data subjects and data controllers, there may be downsides to using them for exercising GDPR rights. As none of these tools are certified by GDPR governing bodies,[78] there is no guarantee that using them will help data subjects better exercise their rights or ensure that data controllers are compliant. Some tools charge fees to data subjects and data controllers for what should be a free right under Article 12(5). Although tools may simplify the process, there are many tools on the market, conversely overcomplicating the process for data controllers and data subjects.

Data Portability within the GDPR

As no empirical work has been done to assess how data portability works in practice, it is unclear how effective the RtDP achieves the GDPR's aim to secure protection for the processing of personal data.

For data subjects as active participants exercising their data protection rights, the RtDP is intended to provide greater agency and control for individuals, particularly for ensuring the free movement of their data. However, there is little assessment of whether this places too much onus on data subjects to right any wrongs when there is discord between different GDPR rights. For example, data subject rights and data protection by design (DPbD) may conflict when exercised and deployed. As a GDPR requirement, DPbD may be restrictive in practice, emphasising privacy-

[77]Port.im (2018) https://www.port.im/ accessed 25 January 2019.
[78]Reg 2016/679 (n 1) art 42.

as-control over privacy-as-confidentiality when no data protection frameworks guide DPbD employment.[79] The privacy-protective Right to Be Forgotten (RtBF) may override the RtDP as a result of 'multiple linking', where two or more data subjects can be easily linked by same datasets.[80] More generally, it has been argued that the GDPR should not be considered as a one-size-fits-all piece of legislation for technology law. Interaction with other regulatory levers and bodies beyond data protection is necessary for a more holistic understanding of regulating such harms.[81] Additionally, collective responsibility may be considered successful for improving the understanding of GDPR responses.[82] De Hert et al. also consider the possibilities for building interoperable infrastructures enabling data subjects to bridge the gap between specific services.[83]

For data controllers, the GDPR rights offered to data subjects may be considered overly expansive, whereby too much emphasis is placed on data subject rights without considering data controllers' and data processors' responsibilities. The content of the responses themselves may include pre-GDPR information, retroactively requiring data controllers to respond. Referring to the then-draft GDPR, Swire and Lagos argue that data portability may reduce consumer welfare as it places excessive burden on small and medium enterprises by disregarding market power and efficiencies.[84] Security challenges may arise as the complexity of controlling and processing

[79]Michael Veale, Reuben Binns, and Jef Ausloos, 'When data protection by design and data subject rights clash' [2018] International Data Privacy Law DOI: 10.1093/idpl/ipy002.
[80]Wenlong Li, 'A tale of two rights: exploring the potential conflict between right to data portability and right to be forgotten under the General Data Protection Regulation' [2018] International Data Privacy Law ipy007 DOI: 10.1093/idpl/ipy007.
[81]Tristan Henderson, 'Does the GDPR help or hinder fair algorithmic decision-making?' (LLM dissertation, Edinburgh Law School 2017) DOI: 10.2139/ssrn.3140887.
[82]Mahieu, Asghari, and Eeten (n 67).
[83]Paul De Hert and others, 'The right to data portability in the GDPR: Towards user-centric interoperability of digital services' (2018) 34(2) Computer Law & Security Review 193 DOI: 10.1016/j.clsr.2017.10.003.
[84]Peter Swire and Yianni Lagos, 'Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique' (2013) 72(2) Maryland Law Review 335 DOI: 10.2139/ssrn.2159157.

personal data increases with more portable data.[85] Further, Graef et al. argue that if justifications for data portability are poorly-defined, portability may be considered a goal in and of itself for data controllers, with little impact on personal data protection.[86]

Although third party tools exist, there are significant privacy and security issues in using these services through the transmission and processing of additional personal data. The sensitive personal data contained in RtDP responses may be inappropriate for sharing with third parties. While such mechanisms may ease the RtDP and RoA process for data subjects, tools may overestimate the difficulty for exercising data subject rights given the large number available as indicated in the previous section. Some private companies, such as Datastreams[87] and Port.im,[88] charge for these tools, no longer making data subject rights free to exercise. By using these tools, the data subject may be providing more of their personal data to new controllers, such as through the Fair & Smart service, which does not clearly state how the collected data is used.[89] As a new GDPR right, empirical work on the implementation of the RtDP is important to see whether the aims of the regulation and its practical application are aligned. Under a technologically neutral framework, it is uncertain how the RtDP should be exercised, particularly with limited guidance for both data controllers and data subjects. While the literature on data portability is well-researched, much of the work pre-dates the GDPR and remains only theoretical in its RtDP application. Empirical research needs to be conducted by exercising the RtDP to assess data portability's normative value within data protection in practice.

---

[85]Stefan Weiss, 'Privacy threat model for data portability in social network applications' (2009) 29(4) International Journal of Information Management 249 DOI: 10.1016/j.ijinfomgt.2009.03.007.
[86]Inge Graef, Martin Husovec, and Nadezhda Purtova, 'Data Portability and Data Control: Lessons for an Emerging Concept in EU Law' [2017] SSRN preprint DOI: 10.2139/ssrn.3071875.
[87]Datastreams (n 71).
[88]Port.im (n 76).
[89]Fair & Smart (n 73).

III. DATA PORTABILITY IN PRACTICE

Methodology

Our aim was to understand how data controllers have approached the RtDP by studying their responses to data portability requests. We focused particularly on the practical application of the law from the perspective of the data subject, without explicitly searching for privacy and data protection considerations. Only the metadata, not the content, of the responses received were assessed. Although the Regulation also offers the right to transmit personal data, we focused only on exercising the right to receive as our research aims to assess the process and responses of RtDP requests with reference to the terms and definitions provided by the GDPR and DPAs.

To exercise the RtDP, we created a Python program to make 230 data portability requests. We used e-mail as the transport for the tool as it is commonly used by both data subjects and data controllers. Although some controllers offer automated download options, this is far from commonplace as indicated by our results below. 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. All interactions with data controllers took place within the EU. To facilitate a more varied set of results, we chose data controllers that spanned a wide range of industries. We categorised these data controllers into 34 categories using the Curlie taxonomy.[90] These categories are demonstrated in Table 1, with the most popular being 'Publications' (12.6%), 'Software' (11.3%), '(Legal) Services' (9.1%), 'Clothing' (6.5%), 'Non-profit Resources' (5.7%), and 'Online Communities' (5.7%).

[90]The Curlie Directory (2019) http://curlie.org/ accessed 25 January 2019.

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 included 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 (DPOs), 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 provided 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.[91] 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 (25 May 2018) and the data collection process ended on the 26 August 2018. The e-mail template and the relevant variables used for the Python program are illustrated and explained in the Supplementary Material.

Results

In this section, we describe how data controllers processed, acted upon, and responded to our RtDP requests. Importantly, and forming the basis for our empirical study, we anticipated that the personal data provided by the data controllers to the data subject would be made 'in a structured, commonly used and machine-readable format'.[92] We analyse the data received and illustrate any patterns we identified.

[91]Information Commissioner's Office, 'Your right of access' (7 June 2018) https://ico.org.uk/your-data-matters/your-right-of-access/ accessed 25 January 2019.
[92]Reg 2016/679 (n 1) art 20(1).

When gathering information for sending out RtDP requests, all data controllers listed a contact e-mail address, where 104 were specifically privacy- or data-protection-related and 126 were general. 173 contact details were found under the terms and conditions or privacy policy pages. However, even after looking at these pages, contact information was sometimes only found after being redirected to another webpage. In 19 cases, the e-mail addresses indicated did not work and an alternative had to be found. The remaining contact information was found through searching the data controllers' webpages.

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 specific e-mail security restrictions. A RtDP request for the latter was resubmitted via its mandated web form. Including responses indicating that no personal data were stored, 172 (74.8%) of the requests were successfully completed, with greater success compared to the Talend study.[93] Five data controllers asked for the full three months allowed under the GDPR, four of which replied within our study period. 25 data controllers responded initially but did not respond to a follow-up e-mail reminder and 33 never responded. Figure 1 demonstrates the response times for the 172 data controllers, represented by a cumulative distribution. 52 (22.6%) replied within a week of the request being made, then 22 (totalling 32.2%) more within two weeks, 67 (totalling 61.3%) within a month, and 31 (totalling 74.8%) within three months. 25% of responses were received by day 7, the median response time was 19 days, and 75% of responses were received by day 29.

[93]Talend (n 70).

1. RtDP request process

In making the requests, 88 data controllers required additional personal data for verifying identity. From these requests, 23 (26.1%) required only an ID, 18 (20.5%) required filling out a designated form, 18 (20.5%) required logging in to personal accounts, 14 (15.9%) required photographic, national ID and proof of address, and two (2.3%) required only a proof of address. All IDs had to be in full view apart from one where the data controller explicitly asked for the ID number to be redacted. Other personal information were asked by 13 (14.8%) controllers, including questions about date of birth, most recent bank transactions, booking references, order numbers, phone numbers, old e-mails, and proof of deactivated e-mail accounts.

When asking for portable data, there was a lack of clarity on GDPR rights. Given the longer existence and wider remit of personal data offered under Article 15, four data controllers misunderstood the RtDP request as a RoA request. Two suggested that we make a RoA request instead so that we can have more personal data. Two RtDP requests were conflated with the RtBF and two with the right to restrict processing.

Although communication began and, for most data controllers, continued through e-mail, the method of receiving RtDP responses varied. 100 (43.5%) responses were sent by e-mail of which 19 (19.0%) were password protected, 20 (8.7%) were downloaded from an online portal of which eight (40.0%) were password protected, 18 (7.8%) were retrieved through personal login accounts, two (0.9%) were file passwords received by post, and two (0.9%) were full postal responses. 32 (13.9%) data controllers indicated that they stored no data beyond the e-mail address and correspondence and did not have additional data to provide. No patterns were identified between the RtDP response time and the data controller category or method of response.

Despite little mention of security in exercising GDPR rights, 16 data controllers explicitly stated in our communication that it shaped what and how personal data were sent and received. Five telephone conversations were required for identification purposes. All data controllers separated the transmission of personal data files and passwords. However, 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 of those used for the RtDP are often flagged by spam filters. One data controller mentioned that they chose XLS files over CSV files because the latter could not be password protected. Notably, one data breach was caused, where an RtDP response included the personal data of other data subjects.

There were uncertainties in how data controllers implemented the RtDP. Four data controllers noted that our RtDP requests were the first that they received. Three data controllers asked for feedback on the process after sending their response. This included questions such as whether the data received were satisfactory, what formats are desirable, whether the communication process was good, and whether the responses were sufficiently timely. One data controller initially said that no data were stored but came back one month later with personal data. Two data controllers explicitly 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.

2. File formats of responses received

Responses were provided in numerous different types (Figure 2), the most popular being tabular CSV or Excel (XLS or XLSX) files. Ten data controllers reported that they chose the file formats (CSV, XML, and JSON) as suggested by the ICO. Other file format categories included PNG screenshots, JPEG images, audio files, e-mail files, and PDF paper scans. The popularity of tabular

formats, representing structured datasets and spreadsheets, broadly indicate how data controllers process data, where raw data is available without executing code.[94] Different file formats may be more suitable for different types of data. For example, where multimedia data were included, documentation types for formatted, page-oriented documents were preferred.

Within categories, there were some patterns in considering the response file type against how data were accessed and received. Figure 3 illustrates the normalised file formats within RtDP responses from data controller categories. The data controller categories displayed in Figure 3 represent over 40% of total formats in total responses by each data controller category. This suggests that certain categories lend themselves to particular file formats, such as Legal Services to Documentation and Publishing to Tabular. It also reflects the kinds of data that is most commonly processed, representing the most appropriate and suitable file formats for such data.

Compliance

Once all data were received, Table 2 identifies the file formats used and whether we thought they were compliant as assessed based on the A29WP's and the ICO's 'structured, commonly used, and machine-readable format' definitions from Section II. Table 2 shows that while compliance is binary (marked YES or NO), it is not always clear where file formats should be placed. Grey areas are marked ?, suggesting certain restrictions required for compliance. For example, CSV files are compliant because there are structural relationships between elements within tabular data, is a

[94]Library of Congress, 'Recommended Formats Statement' (2018) https://www.loc.gov/preservation/resources/rfs/data.html accessed 25 January 2019.

commonly used format, and can be processed by computers. This can be illustrated by a data controller response received that contains a CSV file of previous online purchases with the retailer. The CSV file has clear headers, such as 'product name', 'cost', and 'date of purchase', where the subsequent rows are populated consistently in the same format, for example all dates as YYYY-MM-DD. In contrast, although HTML is a commonly used format for web pages, it is only structured and machine-readable with markup. For example, HTML files that use only content division elements (div) to contain data, although grouped, would not be considered structured. A marked-up HTML file that includes a data structure allowing the computer to access the table data and associate it with a named object, demonstrating a hierarchical relationship between clearly defined variables, would be considered structured. Another file format where compliance may be ambiguous are XLS or XLSX files. Excel spreadsheets may not be machine-readable if multiple sheets are included in a single workbook, where the data may not be automatically processed by a computer. If data is shared across sheets, the relationships between datasets will not automatically be read and identified. Long notes and prose written within a single cell would also render the file format non-compliant as the data cannot be distilled into useful information by a machine.[95] For Excel files to become compliant, a new file should be created for each sheet. Formats may also fall into grey areas where specific technical processing or levels of metadata beyond its default format standard are required. This is often found during the stage where data controllers prepare personal data for transmission to data subjects from their internal devices. Even if the personal data held by data controllers were structured and machine-readable during processing, the extraction process may have made the information conveyed in some files non-compliant. In one response received from a food and related products data controller, the transmission of data via audio files was

[95]Karl W Broman and Kara H Woo, 'Data Organization in Spreadsheets' (2018) 72(1) The American Statistician 2 DOI: 10.1080/00031305.2017.1375989.

acceptable as it was the information provided by the data subject, in this case a telephone call recording that the data subject consented to. It would not be acceptable if it was an audio recording of data that could otherwise have been better represented in another format, such as a CSV file. Similarly with screenshots and paper scans, as was sent by two data controllers, although the information was provided, where data subjects can see what personal data was collected by consent and by contract, the format itself with such data did not adhere to Article 20 requirements. This is because screenshots (such as images of a CSV table) are not machine-readable where the text can no longer be extracted by a computer; paper scans are not machine-readable, even if the information printed was once compiled by a computer. One banking data controller sent paper scans of tables, suggesting that they were printed tabular files that could otherwise have been sent in their original digital form. In another example, a beauty data controller sent paper scans illustrating screenshots of a software interface. In order to the compliant with Article 20 requirements, the data could have been exported into a 'structured, commonly used, and machine-readable' format.

From Table 2, the fully compliant file formats identified based on the definitions provided for 'structured, commonly used, and machine-readable' are CSV, EML, ICS, JSON, MBOX, TEX, VCS, and XML. With a lack of guidance as to what file formats are acceptable in context of certain data types, it is difficult both for data subjects and data controllers to judge what is compliant.

Additionally, for the 20 data controllers that used online 'download your data' portals to provide responses, their RtDP files also used a range of compliant, ambiguously compliant, and non-compliant file formats. Although JSON was the most popular file format, HTML, JPEG, MP4, TXT, and XML were also included. The data controllers who used these tools did not respond quicker, more thoroughly, or with more compliant files overall. Starting the download process itself resulted in quick, often automated, correspondence, but the receipt of RtDP response still took a period of time to be received and available for download. While internal and external tools may be

beneficial for data controllers managing RtDP requests, they may be detrimental for data subjects who must involuntarily provide more information to another third party to exercise their rights.

IV. REVISITING THE RTDP DEFINITIONS FROM A TECHNICAL PERSPECTIVE

From the results in the previous section, we found that RtDP compliance is difficult given the limitations based on technological neutrality, the lack of standards for measuring the appropriateness of file formats, and potential security problems raised by the transmission of personal data. It is clear from the wide range of formats in Table 2 that data controllers are unsure which file formats should be used for RtDP responses. Notably, some of these formats such as PNG screenshots and PDF paper scans are clearly non-compliant. While 82.0% of data controllers responded within a month, this only represents 61.3% of all RtDP requests sent, with the overall completion rate at 74.8% of 230 data controllers. In combination with the ambiguity of file format compliance as indicated in Table 2, we argue that the existing Regulation and guidance materials are insufficient for data controllers to adequately comply with the RtDP.

Through only examining file formats (as opposed to file content), we found that it is difficult to determine what files are compliant and what files are not. From Table 2, it is also unclear to data controllers what ‘structured’ and ‘machine-readable’ entails. The current definitions for ‘structured, commonly used, and machine-readable’ do not provide sufficient guidance for the practical application of data portability. While Article 20 defines the requirements for compliance as discussed in Section II, technologically neutral language limits the usefulness of practical RtDP application. It should be acknowledged that maintaining technological neutrality does not equal not using technical language or considering technology in applying the law. Hildebrandt and

Tielemans identify the technology neutral law aims to meet certain compensation, innovation, and sustainability objectives.[96] In order to meet these objectives, normative dimensions with the attention and discernment of the legislator that go beyond black-and-white legislation are required. The authors go further to suggest that technology specific legislation is needed to ensure that the objectives of technology neutral law are met. While the existing A29WP Guidance explained what data needs to or should be provided, those definitions are inadequate in practice.

In supporting compliance, we attempt to redefine RtDP terms by suggesting technical definitions to achieve the GDPR's goals. From the data gathered, we consider new data portability definitions, clarify how data should be made portable, and explain the appropriateness of file formats in relation to how data could be determined according to their type or industry. Both directed at data controllers, setting new data portability definitions could support the exercise and compliance of the RtDP. Whether a file format is appropriate for certain RtDP responses is contextual. In order to determine the best format for certain portable data, it is necessary to assess the metadata returned 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. From one of the responses received from a Clothing data controller, we received a RtDP compliant CSV file clearly stating that it was of previous financial transactions. The CSV file was labelled with headers that were consistently defined such as 'date (YYYY-MM-DD)', 'transaction item (ID)', and 'payment details (card ending XXX)', and 'amount (in £)'. From an Online Communities data controller, we received multiple JSON files within our RtDP response, where each file name was a separate category of information such as 'account', 'friends', and 'search history'. Each file contained data about those categories, again with consistently defined headers. More detailed guidance on file formats alone,

[96]Mireille Hildebrandt and Laura Tielemans, 'Data protection by design and technology neutral law' (2013) 29(5) Computer Law & Security Review 509 DOI: 10.1016/j.clsr.2013.07.004.

without identification of data management or best practices that relate to those file formats, are inadequate for ensuring data portability. For data subjects, having further guidance on how they can know and check the definitions regarding the appropriate decoding of data file formats, file data content in context of file formats, and data organisation practices should be elaborated upon. This information could be provided by the data controller or the relevant DPA. Two data controllers' RtDP responses demonstrated how they believed that their files were RtDP compliant by including an index HTML file that explained the structure of the JSON file, what each of the objects mean, and what the data represents by also representing their 'structured, commonly used, and machine-readable' files in a human-readable format.

Structured

The Article 20 definition of 'structured' fails to represent the necessity of metadata in exercising data portability and only goes so far as requiring a human-readable label within unstructured text. While the A29WP indicated that data controllers should provide 'as many metadata with the data as possible at the best possible level of granularity'[97], the lack of detail on what type, suggested headings, or potential standardisations for clearly defined metadata make it difficult for data controllers to implement the 'structured' requirement in practice. In Computer Science, 'structured' refers to a high-level of organisation where data can be searched and queried within hierarchical orders. What makes structured data 'structured' is its well-defined, logical representation where structural relations between elements are explicit.[98] The example used by the ICO to illustrate a structured format, a spreadsheet, represents a very specific type of structured data. An example of a well-structured file from an Associations data controller's included multiple CSV files. Each

[97]Article 29 Data Protection Working Party (n 19) 14.
[98]Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Pearson Education Limited 2016) 58.

file's name indicated what the data content of the file would be such as 'administrative e-mail codes', 'support messages', and 'login history'. Each CSV file included headers that described the metadata for each column of data. The high-level of properly arranged metadata ensure that the spreadsheets are structured, relational database systems. However, without a clear hierarchical, relationship between values or classification, metadata is mapped only to a label without co-dependence. As a result, from Table 2, we identified CSV, EML, ICS, JSON, KMZ, MBOX, TEX, VCS, XLS/XLSX, and XML file formats as structured, where data is relational within a well-defined mathematical structure for access and manipulation when the structural relations within the metadata is explicitly defined within those files.

Instead of a binary consideration for structured or non-structured data, the idea of semi-structured data should be introduced in the RtDP to better reflect digital data in practice. By nature, much of the personal data processed go beyond full-text documents and databases, suitably defined as 'structured', where information can no longer be constrained by a schema. This type of data without a rigid data model is referred to as semi-structured.[99] The GDPR does not explicitly differentiate between structured and semi-structured data. Given the nature of personal data that falls under GDPR requirements, inclusion of semi-structured data is implied. However, if that is the case, there are significant limitations to the existing definition of 'structured'. Although JSON and XML, both cited by the ICO as acceptable for the RtDP, are common semi-structured formats, the relatively lax constraints diminishes the ability for data reuse if there is not enough metadata. An example of semi-structured data, and one received by a Fitness data controller, is an X-ray. While the content of X-rays cannot be searched as if it were a structured database, the scan's

[99]Peter Buneman, 'Semistructured Data' (Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Tucson, 1997) DOI: 10.1145/263661.263675.

metadata could include useful information such as when it was created, who it was created by, and whose body the image is of.

Requirements for semi-structured and structured data can be made more specific without detriment to technological neutrality within the GDPR. Under the RtDP, 'structured' should clarify the importance of metadata. Without more robust requirements on file format structures, the success of data portability as a means for data subjects' ability to transmit their personal data for the purposes of data protection is undermined.

Commonly Used

While no further clarification is provided beyond 'widely-used and well-established', our empirical results demonstrate that there is agreement on 'commonly used' and is generally settled in application. At present, a 'commonly-used' file format could be considered as a format where, using the terms from GDPR Article 20(2), it is 'technically feasible' for it to be 'machine-readable' without adopting new technologies. Data subjects should also be reasonably expected to be able to access the RtDP response files on a popular operating system such as Linux, MacOS, or Windows. It may be possible for 'commonly used' formats to be pre-approved by standards bodies to ensure that any new file formats comply with existing data portability or industry-based standards. For example, the only file format we received that was not 'commonly-used' was a KMZ file for mapping data. As a zipped KML file, primarily used by Google Maps, the KMZ may not uncompress on all geobrowsers. As a result, a more commonly-used file format should be KML files, the standard agreed upon by the Open Geospatial Consortium.[100]

[100]Open Geospatial Consortium (2014) http://www.opengeospatial.org/standards/kml/ accessed 25 January 2019.

Machine-Readable

Given that the ICO describes most structured data as machine-readable, it suggests that 'machine-readable' refers to the application of portable data. Beyond the A29WP, the ICO has a non-codified minimum requirement for the extent to which machine-readability should be open, using the example of Excel files as less optimal because of its proprietary and encrypted nature. Machine-readable can be distinguished between markedup human-readable data and files that are intended primarily for (further) machine processing. While some formats such as XML and XSLT are both human- and machine-readable, data that is initially human-readable does not automatically become machine-readable when transferred to a machine-readable format. Data that is represented and formatted in a Word DOCX file and is structured to the human eye does not necessarily mean that it is readable by a machine as information from the data cannot be extracted for further processing. Without clarification on what machine-readability entails, even the file formats suggested by the ICO (CSV, JSON, and XML) may not necessarily compliant all the time if they are not as well-defined as they could be. For example, CSV files without headers are not self-documenting and if the null values are handled differently within the file, the CSV becomes less machine-readable, making the contained information less useful. Certain existing technologies could be recommended for data controllers for compliance. For example, for paper scans, many scanners have an Optical Character Recognition (OCR) function, converting images to text and PDF to text.

Rather than defining 'machine-readable' only as automatically readable and processed by a computer, requirements should be created as a measure for machine-readability. File formats on their own can be machine-readable, but without sufficient metadata, the content cannot be processed further. For example, Data.gov lists the factors of uniformity, simplicity, ubiquity,

economy, and extensibility as conditions to ensure increased data capabilities.[101] While there is a danger that the creation of standards may be overbearing and messy, a network of standards can support a world enacted through technical practicalities for better construction of our realities in the ever-evolving digital ecosystem that process personal data.[102] Regulated standard setting is already part of EU operations. Regulation 1025/2012 Annex II illustrates how the European Commission's standards may be met, suggesting bodies such as the W3C or Internet Engineering Task Force (IETF) for doing so.[103] Sector standards may be set to shift from compatible to interoperable systems.

Summary

Technically-advanced definitions for 'structured, commonly used, and machine-readable' could be provided so that it becomes actionable, allowing appropriate tests to be designed without neglecting GDPR technological neutrality. Further guidance on the RtDP must include technologically sensitive definitions of 'structured, commonly used, and machine-readable format' that addresses semi-structured data processed under Internet systems and the practicalities of extracting portable data into transmittable files. Personal data under the RtDP should not strictly be defined as structured or unstructured, but should include semi-structured data. While 'commonly used' is sufficiently clear in practice, independent or industry bodies could be established to assess the commonality to certain file formats. As 'machine-readable' is already standardised within certain

[101]Datagov, 'A Primer on Machine Readability for Online Documents and Data' (2018) https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data accessed 25 January 2019.
[102]Lawrence Busch, *Standards: Recipes for Reality* (1st edn, MIT Press 2011).
[103]Regulation (EU) 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation [2012] OJ L316/12.

EU industries, a European-wide definition for ‘machine-readable’ should be adopted and further clarified in relation to the RtDP for the benefit of both data controllers and data subjects.

New RtDP definitions of ‘structured, commonly used, and machine-readable’ formats could be adopted without relying on other legal mechanisms for better data protection.

V. FUTURE WORK

Our results show problems around portability both for data controllers, who may misunderstand RtDP requirements or provide data in inappropriate or incomplete formats, and data subjects, who may be unable to verify their identity or veracity of the data returned by a controller. While our research aims to present a broad range of data controllers, it may be argued that the scope was not wide enough to adequately represent the treatment of the RtDP. Too little data may be gathered from the responses received without gaining a better understanding of the data controllers’ internal process for responding to RtDP requests. As the requests were conducted shortly after the GDPR came into effect, it may be too early to tell how data controllers aim to fully comply. In analysing the results, the time spent on services was not normalised. As a greater time spent on data controllers’ platforms may result in a wider range of data and file types to be collected, this may have affected the types of personal data supplied by data controllers. Our study is still at an early stage and raises several possibilities for future work.

Beyond establishing new definitions for exercising the RtDP, new guidance, standards or codes, further empirical research, and the development of technological solutions can help clarify the right for all stakeholders. Collaboration between lawyers, policy-makers, enforcement bodies, data controllers, and technologists is highly encouraged to ensure that data portability is viable in theory and in practice.

New Guidance, Standards, or Codes of Conduct

From revisiting the RtDP definitions, we attempted to expand the technical definitions for 'structured, commonly used, and machine-readable' file formats in an attempt to help data controllers comply with the RtDP. Although the patterns identified in Figure 3 may not be significant or indicate strong preferences for certain file types by specific categories, it suggests that certain processes can be streamlined by establishing standards or best practices in making data portable. This is particularly useful given that, based on the data controller categories and file formats received, we were unable to find any industry standards for data portability. In addition to the UK ICO's guide, it may be useful to assess any additional, supplementary guidance documents produced by other EU countries' DPAs to see the similarities and differences in adopting the RtDP. Beyond considering new technical definitions, future work in this area could include the development of new guidance, standards, or codes of conduct.

As the GDPR becomes a more mature piece of legislation, it may be possible that new legal decisions will encourage the EDPB and DPAs as enforcers of the legislation to produce new guidance materials. For instance, in competition law, TFEU Article 102 is supplemented by the European Commission's 2005 Discussion Paper[104] and 2009 Guidance Paper[105] amongst many non-binding AG Opinions such as *The Scotch Whisky Association*,[106] *Intel Corporation Inc.*,[107] and *Coty Germany GmbH*.[108] As discussed in Section II, data portability was grounded in competition

[104]European Commission, 'DG Competition discussion paper on the application of Article 82 of the Treaty to exclusionary abuses' (2005) 1 Official Journal of the European Union 72.
[105]European Commission, 'Communication from the Commission – Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings' (2009) C45 Official Journal of the European Union 7.
[106]'Scotch Whisky Association and Others v The Lord Advocate and The Advocate General for Scotland' [2015] Case C-333/14 Opinion of Advocate General Bot.
[107]'Intel Corporation Inc v European Commission' [2016] Case C-413/14 P Opinion of Advocate General Wahl.
[108]'Coty Germany GmbH v Parfümerie Akzente GmbH' [2017] Case C-230/16 Opinion of Advocate General Wahl.

law pre-GDPR, and so there may be scope for a wider range of guidance materials to be provided in support of data portability enforcement. Without new legal decisions, however, only time will tell as to whether the EDPB and other DPAs will issue further guidance, developing new standards and legal tests.

Following GDPR Article 40, associations and other bodies representing categories of controllers could work together and create new codes of conduct for responding to RtDP requests. This would not only support data controllers in complying with the RtDP but also help data subjects exercise their right. For instance, the EU Cloud Select Industry Group (C-SIG) and A29WP has established a EU Data Protection Code of Conduct for Cloud Service Providers[109] that takes into consideration GDPR requirements and represents a good starting point for the consideration of codes of conduct within data protection regulation. The scope of this Code is restricted to Cloud data processors, however, and only deals with data regulation more generally. References to data portability was only introduced by C-SIG after taking the A29WP Opinion[110] into account. Even then, it is only briefly mentioned in the Code within its consideration of data subject rights.[111] While the C-SIG Code would not be directly applicable to implementing the RtDP, it demonstrates the potential for establishing a code of conduct for Article 20, where each recommendation within the code could align data controller responsibilities with a means to address an issue within the RtDP that extend beyond references to the Regulation.

To find out whether guidance, standards, or enforceable codes of conduct would work best in the context of the RtDP, lawyers, policy makers, computer scientists, and data controllers should

[109]EU Cloud Select Industry Group, 'EU Data Protection Code of Conduct for Cloud Service Providers' (v2.1, 2018).
[110] Article 29 Working Party, 'Opinion 02/2015 on C-SIG Code of Conduct on Cloud Computing' (WP232, 2015) 12.
[111] EU Cloud Select Industry Group (n 108) 22.

work together to contextualise portable data within a practical data protection framework that succeeds in achieving the aims of the GDPR.

Empirical Research

One avenue for future work is by looking at the content of the responses received. Due to the scope of our research, the RtDP content responses were not assessed, limiting how compliance may be interpreted, particularly with respect to whether the data contained within certain file formats were the most appropriate. By examining the content of responses and determining whether the desired information could be extracted through machine-readable means, file format standards can be set for corresponding data.

As our RtDP requests were made within the UK, the exercise of the RtDP could be replicated in other countries both governed by the GDPR and in other jurisdictions to assess whether the responses received are similar. This would allow a more comprehensive view of the right. Any best practices or guidance from local data protection authorities could be adopted and developed into a EU-wide framework to support making data portable.

In our research, we examined data transfers from data controller to data subject, but Article 20(3) 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.[112] The interoperability of specific categories, such as social networks, can be explored based on the metadata received from RtDP responses. 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.

Technological Solutions

Finally, technological routes can both support data subjects to exercise data protection rights, such as the RtDP, and provide data controllers with secure identification and storage tools to make it easier for verifying requests.

For data subjects, edge computing may improve scalability, responsiveness, and privacy policy enforcement for data subjects.[113] The module could make RtDP requests on behalf of data subjects, securely store the responses, aid verification, and sanitise responses to be shared with other data subjects to crowd-source quality information about data controllers for compliance. Chatbots may support exercising the RtDP where websites are automatically parsed to obtain data controller

[112]Article 29 Data Protection Working Party (n 19) 11.

[113]Mahadev Satyanarayanan, 'The Emergence of Edge Computing' (2017) 50(1) Computer 30 DOI: 10.1109/mc.2017.9.

contact details, providing data subjects with user-friendly information.[114] The GDPR Chatbot[115] and Parker[116] already assist in explaining GDPR requirements.

For data controllers, open application programming interfaces (APIs) may be encouraged across certain data controller industries, where proxies translate data formats and semantics. Use of REST API functions and libraries can support interconnected open-source Personal Data Store servers that allow different applications on different devices to read and write data .[117] New open standards for Decentralised Identifiers and Authentication ,[118] Decentralised Key Management System,[119] and Verifiable Credentials[120] can strengthen the self-sovereign identity (SSI), where peer-to-peer relationships between people and organisations may preserve the data subject's credentials and relationships securely.[121]

Third parties could support the RtDP by introducing new frameworks and technologies for data protection, such as the aforementioned Data Transfer Project.[122] There is, on the other hand, the danger that these may lead to the adoption of self-regulatory models of data protection enforcement. Mechanisms to engage more stakeholders could help here. For instance, the GDPR's regulatory levers such as codes of conduct could be employed, with codes of conducts to be established by EDPB-recognised third party groups for data portability that involve a wide range

---

[114]Hamza Harkous and others, 'PriBots: Conversational Privacy with Chatbots' (Twelfth Symposium on Usable Privacy and Security,Denver, 2016).
[115]GDPR Chatbot (2019) https://gdpr-chatbot.com/ accessed 25 January 2019.
[116]Norton Rose Fulbright (14 May 2018) http://www.nortonrosefulbright.com/news/167374/norton-rose-fulbright-launches-ai-powered-chatbot-on-eu-gdpr accessed 25 January 2019.
[117]Personium, 'Documentation' (2018) https://personium.io/docs/en/ accessed 25 January 2019.
[118]W3C Community Group, 'Decentralized Identifiers (DIDs)' (22 January 2019) https://w3c-ccg.github.io/did-spec/ accessed 25 January 2019.
[119]Michael Egorov, MacLane Wilkison, and David Nuñez, 'NuCypher KMS: Decentralized key management system' (15 November 2017) https://arxiv.org/pdf/1707.06140.pdf accessed 25 January 2019.
[120]IMS Global Open Badges Workgroup and W3C Verifiable Credentials Working Group, 'Drafts and Ideas of Educational and Occupational Verifiable Credentials' (19 January 2019) https://github.com/w3c-ccg/eduoccverifiablecredentials accessed 25 January 2019.
[121]Sovrin, 'Sovrin: A Protocol and Token for Self-Sovereign Identity and Decentralized Trust' (January 2018) https://sovrin.org/wp-content/uploads/2018/03/Sovrin-Protocol-and-Token-White-Paper.pdf accessed 25 January 2019.
[122]Data Transfer Project (n 72).

of stakeholders. Technology could also involve multiple stakeholders; for instance trusted differential privacy focused data repositories could be collectively maintained with crowdsourced metadata for building a better understanding of the RtDP.[123]

In summary, investments into technological solutions such as edge computing, chatbots, and differential privacy research, may be able to fill technical gaps in the GDPR resulting from the Regulation's technological neutrality and help mandate better security measures.

VI. CONCLUSION

In this paper we exercised the GDPR's RtDP by making 230 real-world requests and examined the responses from data controllers. As a new data subject right that requires certain technologies for its implementation, we found that the GDPR's technological neutrality restricted the Regulation's ability to sufficiently explain how the RtDP should be exercised and implemented. From our results, we found the process of making RtDP requests to be cumbersome, where e-mails were not always easy to find and some data controllers were unresponsive. In the responses received, we found a variety of file formats being returned by data controllers, some of which may not comply with the RtDP obligations, and some confusion between the various rights in the GDPR on the part of data controllers. Based on our observations, we revisit the definitions for 'structured, commonly used, and machine-readable' RtDP file formats, expanding current definitions with technical details for the purposes of supporting data controller compliance. Future work is needed to help both data controllers and data subjects understand and exercise the RtDP. We suggest that various stakeholders work together to decide the most appropriate method for supporting the RtDP whether that may be the development of guidance, standards, or codes of conduct. Further empirical

[123]Cynthia Dwork, 'Differential Privacy: A Survey of Results' (Theory and Applications of Models of Computation–TAMC, 2008) vol 4978 DOI: 10.1007/978-3-540-79228-4˙1.

research, particularly in other countries bound by the GDPR, would provide a more holistic view of the RtDP in practice. Finally, technological solutions that maintain the technologically neutral GDPR framework are considered, whereby new tools can be built to support the existing Regulation.

TABLES AND FIGURES

Data Controller Category	Number of Data Controllers
Publications	29 (12.6%)
Software	26 (11.3%)
(Legal) Services	21 (9.1%)
Clothing	15 (6.5%)
Non-profit Resources	13 (5.7%)
Online Communities	13 (5.7%)
Publishing	11 (4.8%)
Travel	11 (4.8%)
Event Planning and Production	10 (4.3%)
Arts and Entertainment	7 (3.0%)
Food and Related Products	7 (3.0%)
Attractions	5 (2.2%)
Fundraising	5 (2.2%)
Government	5 (2.2%)
Associations	4 (1.7%)
Beauty	4 (1.7%)
Distance Learning	4 (1.7%)
Education	4 (1.7%)
Financial Services	4 (1.7%)
Human Resources (Management)	4 (1.7%)
Stationary	4 (1.7%)
E-commerce	3 (1.3%)
Hospitality	3 (1.3%)
Telecommunications	3 (1.3%)
Accounting	2 (0.9%)
Fitness	2 (0.9%)
Marketing	2 (0.9%)
Real Estate	2 (0.9%)
Transportation and Logistics	2 (0.9%)
Design	1 (0.4%)
Engineering	1 (0.4%)
Hardware	1 (0.4%)
Maintenance	1 (0.4%)
Unions	1 (0.4%)

Table 1: Data controllers organised by their categories according to the Curlie taxonomy.

File Format	Structured?	Commonly Used?	Machine-Readable?
Email body	NO	YES	NO
CSV	YES	YES	YES
DOC/DOCX	NO	YES	NO
EML	YES	YES	YES
HTML	?	YES	?
ICS	YES	YES	YES
JPEG	NO	YES	NO
JSON	YES	YES	YES
KMZ	YES	NO	YES
MBOX	YES	YES	YES
MP4	NO	YES	NO
PDF	?	YES	?
PNG	NO	YES	NO
RTF	NO	YES	NO
TEX	YES	YES	YES
TXT	?	YES	?
VCS	YES	YES	YES
WAV	NO	YES	NO
XLS/XLSX	YES	YES	?
XML	YES	YES	YES

Table 2: RtDP file format compliance based on 'structured, commonly used, and machine-readable' requirements. Most formats either comply (YES) or do not comply (NO) with the ICO definitions but some are ambiguous (?).

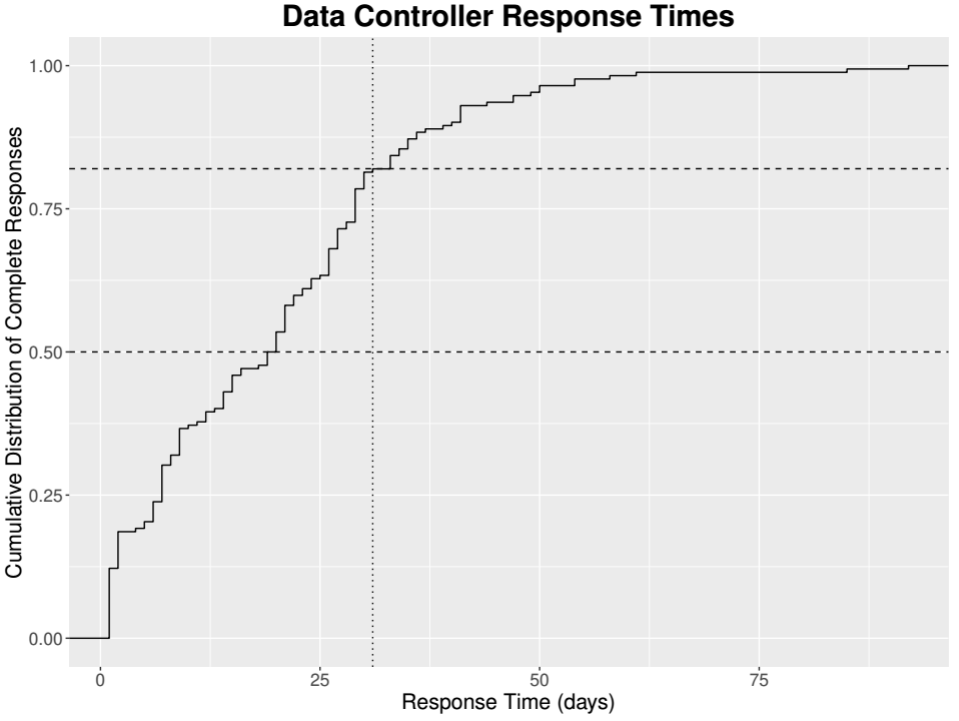

Figure 1: Data controller response times for the 172 complete RtDP responses. The vertical dotted line indicates the required one month (31 days) mark. The horizontal dashed lines indicate the number of responses at 50% and 82% where the line intersects with the one month mark. While 172 data controllers responded to RtDP requests, only 82% of responses (63.5% of the total 230) responded within one month.

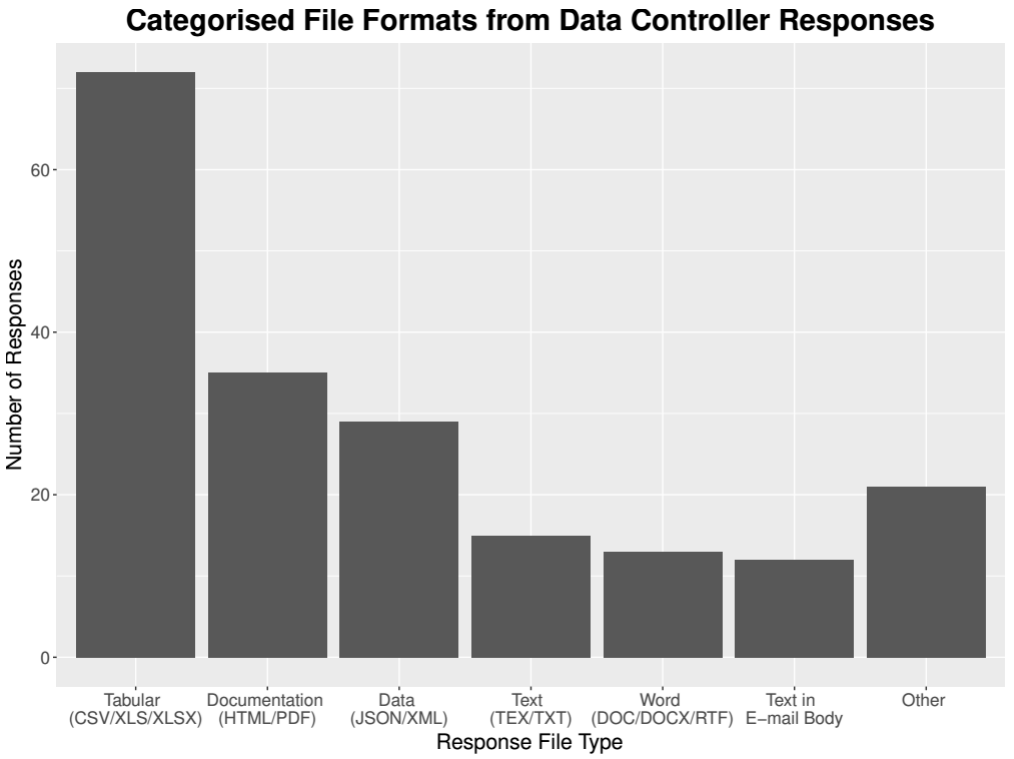

Figure 2: Categorised file formats from data controller responses. The popularity of tabular formats, representing structured datasets and spreadsheets, broadly indicate how data controllers process data, representing 36.5% of all normalised response file types received.

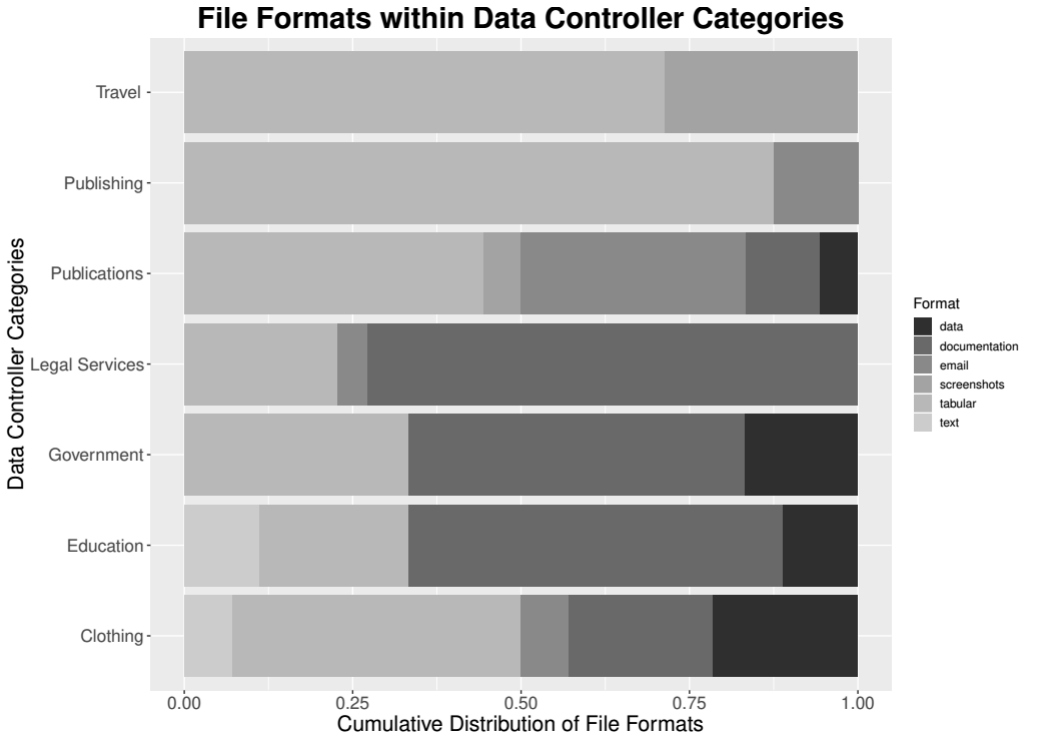

Figure 3: Normalised file formats within data controller categories RtDP responses. The data controller categories shown have at least one file format with over a distribution of 0.4 represented in the responses, demonstrating the formats preference for certain file formats within data controller categories.