DATA PORTABILITY

BY DANIEL L. RUBINFELD[1]

CPI ANTITRUST CHRONICLE
NOVEMBER 2020

Visit **www.competitionpolicyinternational.com** for access to these articles and more!

I. INTRODUCTION

Data is an essential raw material in our economy. Predictions based on relationships identified in data affect numerous aspects of our lives, yet much of this data is collected in a world that is largely modular.[2] To illustrate, it is predicted that by 2025 seventy-five billion Internet of Things devices, controlled by numerous market players, will be connected to the internet, collecting and using data.[3] Furthermore, for those concerned about privacy, data lie at the core of personal profiling in all of our social networking sites.

In each and every case, data portability (the ability to transfer data without affecting its content) and interoperability (the ability to integrate two or more datasets) significantly affect the use of data, with important implications for antitrust policy. Currently, it can be difficult for users to move personal data in a social network to a competing service.[4] Allowing for improved data portability could facilitate the ability of consumers to switch services, which would substantially increase competition. To illustrate, a Facebook user could readily connect with users of other social networks, irrespective of their initial social network provider. In addition, data interoperability can create data synergies: combining data from different sources can improve the knowledge that can be mined from it.

Barriers to data portability and the interoperability that greatly increases the private benefits of portability can be a major source of social inefficiency in our data-intensive economy.[5] In response to concerns of this type, the European Commission has promulgated its General Data Protection Regulation ("GDPR"). Put simply, the GDPR puts into place a regulatory overlay under which the Competition Directorate and the member states can manage competition policy.[6]

---

2 GREG ALLEN & TANIEL CHAN, ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY 27 (2017).

3 *Internet of Things (IoT) Connected Devices Installed Base Worldwide from 2015 to 2025 (in Billions)*, STATISTA, https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/ (last visited November 1, 2020).

4 But, Google's Data Transfer Project has made the movement of photos and other personal data easier. See https://github.com/google/data-transfer-project (accessed on November 7, 2020). According to the Project, "We are establishing a common framework, including data models and protocols, to enable direct transfer of data both into and out of participating online service providers."

5 Oscar Borgogno & Giuseppe Colangelo, *Data Sharing and Interoperability: Fostering Competition Through APIs*, 35 COMPUT. L. & SEC. REV. (2019).

6 Commission Regulation 2016/679, 2016 O.J. (L119) Art. 18 ¶ 2 [hereinafter GDPR]. See European Commission, *A Digital Single Market Strategy for Europe*, 14-15(2015). See also JAN KRAMER, PIERRE SENELLART & ALEXANDRE DE STREEL, MAKING DATA PORTABILITY MORE EFFECTIVE FOR THE DIGITAL ECONOMY: ECONOMIC IMPLICATIONS AND REGULATORY CHALLENGES (2020).

To this point, the U.S. has yet to follow suit. At the core of data-related concerns in the Biden administration are likely to be interoperability and portability.[7] Questions abound. Does imposing a regulatory overlay, perhaps modeled on our telecom regulation, make sense? Short of creating a new federal agency, can the FTC utilize its rule-making functionality to achieve substantial *ex ante* regulatory-like benefits? Or, can the DOJ achieve similar benefits through litigation and/or the use of its typically *ex post* consent decree power? Will active antitrust enforcement improve efficiency, or will it lead to remedies that stifle innovation?

There is a real concern that barriers to data sharing could result in the balkanization of data within particular sectors or even firms, thereby not only impeding innovation within markets, but also reducing spillovers to other markets. Indeed, it is quite possible, when all is considered, that private concerns and private regulation could prevent the sharing of data that would otherwise be efficiency-enhancing.[8]

The discussion that follows lays out some of the pros and cons of a move towards requirements of data interoperability and portability, whether through regulation or through antitrust enforcement.[9] There are technological obstacles to widening the use of data that can be overcome through data portability. Whether the push for interoperability and portability will require a more interventionist role for our competition authorities in order to deal with those obstacles is an open question. On the plus side, standardizing data so that they are portable can lead to smoother data flows, better machine learning, and easier policing of infringement, and can reduce any adverse effects of data-fed algorithms. Standardization might also support a more competitive and distributed data collection ecosystem. At the same time, increasing the scale and scope of data analysis can create negative externalities in the form of better profiling, increased harms to privacy, and cybersecurity concerns.

II. DATA: ANALYSIS AND MARKETS

To understand interoperability and portability issues, it is important to explore the relevant characteristics of data, data analysis, and data markets, as well as some technological obstacles to the use of data and to data integration. While some types of data are not fungible,[10] other datasets can be relevant for multiple users, operating in a wide variety of markets.[11] Moreover, many of those markets are two-sided, as for example the market for Google search advertising, with consumers on one side and advertisers on the other.[12] Furthermore, big data has increased the ability of algorithms to reveal interesting relationships between attributes of datasets and to mine valuable knowledge for descriptive as well as predictive functions.

In an ideal world, data would be transferable or replicable at very low marginal cost. In principle, interoperability can be achieved because data are divisible and can potentially be integrated with other data. Moreover, when economies of scaleand scope cannot be achieved by a single entity or by a single source of data, data integration has the potential to significantly increase data's predictive value. While obstacles abound, in some cases portability will be essential if the benefits of the integration of large amounts of data into one high-quality dataset can be achieved. The challenge is to integrate data that are not necessarily similar in source or structure and to do so quickly and at a reasonable cost.[13]

---

7 The European Commission has seen APIs (Application Programming Interfaces) as vital to achieving interoperability and through portability to make possible the flourishing of Artificial Intelligence and the Internet of Things. See Borgogno &Colangelo, *supra* note 5, at 4 (noting that the GDPR (Article 20) envisions a series of data portability rights that will support the free-flow of non-personal data and the re-use of government data). The authors stress that data sharing through APIs requires a complex implementation process and standardization for success. Similarly, Article 6 of the GDPR creates a right to business-to-business data portability. For further background, see, e.g. Orla Lynskey, *Aligning Data Protection Rights with Competition Law Remedies? The GDPR Right to Data Portability*, 42 Eur. L. Rev. 793 (2017). See also Jorg Hoffman & Begona Gonzelez Otero, *Demystifying the Role of Data Interoperability in the Access and Sharing Debate* (Max Planck Institute for Innovation & Competition Research Paper No. 2016, 2020).

8 See Catherine Tucker, *Online Advertising and Antitrust: Network Effects, Switching Costs, and Data as an Essential Facility*, CPI Antitrust Chronicle (April 2019). See also Aysem Diker Vanberg & Mehmet B. Unver, *The Right to Data Portability in the GDPR and EU Competition Law: Odd Couple or Dynamic Duo*? 8 Eur. J. L. & Tech. 1 (2017) (suggesting that lessons can be learned from EU competition law to limit the potential adverse consequences of the right to data portability).

9 For a more extensive discussion of the benefits and costs of data standardization that is essential for portability to be effective, see Michal Gal & Daniel L Rubinfeld, *Data Standardization*, 94 N.Y.U. Law Rev. 737 (2019).

10 Maurice Stucke & Alan Grunes, Big Data and Competition Policy (2016).

11 See Anja Lambrecht & Catherine E. Tucker, *Can Big Data Protect a Firm from Competition*? Competition Policy Int'l (2017).

12 James Ratliff & Daniel L. Rubinfeld, *Is There a Market for Organic Search Engine Results and Can Their Manipulation Give Rise to Antitrust Liability*, 10 J. Competition L. & Econ. 517 (2014).

13 *The 6 Challenges of Big Data Integration*, Flydata, https://www.flydata.com/the-6-challenges-of-big-data-integration/ (last visited November 3, 2020).

Competition for data collection, analysis, and storage, as well as competition in markets for data-based products or services, is shaped by the height of entry barriers at various points in the vertical chain, from manufacturer to wholesaler-distributor to retailer and finally to the consumer. The demand for data has created an ecosystem of numerous firms that trade in data.[14] This, in turn, enables firms to use data collected elsewhere to scale up their datasets.

Currently, a number of collaborative projects directed towards improving data interoperability and portability are underway. Founded by Google, Microsoft, Yahoo, and Yandex, schema.org is a collaborative effort to create, maintain, and promote schemas for structured data – in essence, to achieve data standardization. Projects include "dataset search" and "the data commons." Built by schema.org, datacommons.org is an open knowledge repository that combines data from public datasets using mapped common entities. In addition, Google Takeout allows users to export their data in an "industry standard" form.[15]

There are three obstacles to achieving substantial benefits from portability. The first involves *metadata uncertainties*.[16] Metadata comprise the data that describe the data included in a dataset. Metadata uncertainties limit others' ability to understand what different data points signify (e.g. does the label "address" relate to billing or to shipping). As such, metadata can increase information asymmetries regarding the content of datasets, thereby reducing incentives to engage in mutually beneficial data sharing.

The second limitation involves *obstacles to data transformation*, which can raise the costs of combining the available data into coherent datasets i.e. achieving data interoperability. One such obstacle results from data granularity, as when similarly attributed data are collected at different times. Another obstacle can arise from the need to reorganize data into a new, combined dataset with a different structure or internal organization.

The third obstacle involves *missing data*. This limitation, which is difficult to correct ex post, arises when some necessary data were not collected, or the costs of ex post collection is prohibitive. Missing data may also result, for example, from limited capacity of a database to store the data,[17] or from data collectors' limited foreseeability of the value of data interoperability.

These three limitations reduce users' incentives and ability to extend the use of data and to achieve data synergies. Indeed, a European Commission study found that "merging different datasets and making them interoperable is one of the most resource-intensive activities for data (re-)users and that, even within the same value chain, datasets are rarely interoperable by default."[18]

III. THE BENEFITS AND COSTS OF PORTABILITY

Making data interoperable and portable can potentially reduce all of the obstacles to data use by others. At the same time, creating functional standards can increase the potential value of portability. Data standardization can increase interoperability (of datasets), lower switching costs for consumers (from one data collector to another), and limit duplication (of data collection, storage and analysis). The threat or actuality of increased competition can reduce the market power of economically powerful platforms, lowering prices paid, directly or indirectly, by consumers.

Supporting the interoperability of different data sources also reduces investment risks associated with data collection, organization and storage. By reducing data portability costs and enabling more market players to utilize data, data standardization may increase incentives for data sharing. Increased use of data may also facilitate cumulative and synergetic knowledge production.[19]

---

14 U.S. Senate Comm. on Commerce, Sci., & Transp., Off. of Oversight & Investigations, Majority Staff, A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes 20 (Dec. 18, 2013), http://educationnewyork.com/files/rockefeller_databroker.pdf.

15 See, for example, www.lifewire.com/what-is-google-takeout-4173795.

16 Avigdor Gal, Uncertain Schema Matching (2011).

17 For an in-depth analysis of the problems involved in collecting and storing data, see, e.g. Blue Ribbon Task Force on Sustainable Dig. Preservation & Access, Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information (2010), https://www.cs.rpi.edu/~bermaf/BRTF_Final_Report.pdf.

18 Eur. Comm'n, *supra* note 6, at 89.

19 This was recognized by the European Commission: *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Towards a Thriving Data-Driven Economy* 14-15, COM (2015). Realizing potential data synergies also depends on the information market participants possess regarding relevant datasets. See Barbara Engels, *Data Portability Among Online Platforms*, 5 Internet Policy Rev. 1, 9 (2016).

www.competitionpolicyinternational.com
Competition Policy International, Inc. 2020© Copying, reprinting, or distributing
this article is forbidden by anyone other than the publisher or author.

Data portability can support a competitive and distributed data collection ecosystem. Not only can it increase the incentives of firms to collect and to share data, it can make markets more competitive. It can also increase the ability of firms to integrate different datasets and reduce the need to rely on one source for data, either internal or external. For example, Google may combine data regarding a user's email, geo-location, and browser history, to better predict her preferences. Other firms, which lack such a variety of data sources, may find it difficult to match these capabilities.

The quality gap created by such network effects carries the potential to entrench or strengthen the dominance of some firms. As a result, data-based markets could exhibit highly concentrated structures, with a single dominant firm possessing a massive share. Benefits arising from data collection and analysis that are not the result of artificial entry barriers do not in themselves raise antitrust issues.[20] However, some have advocated the need for a regulatory overlay.[21]

The difficulty of achieving scale may be overcome if competitors could combine data collected by numerous sources. The lower the costs and obstacles to data portability and interoperability, the stronger the potential competitive pressures on large data collectors. And, since data are non-rivalrous and often easily replicable, data collectors could share their data with many potential users, potentially strengthening competition even further. But the potential has yet to be fully achieved. To illustrate, the recipients of data obtained from Google Takeout may use a variety of different industry standards.

Other barriers, such as switching costs, may still exist.[22] Moreover, a more dispersed market structure might come with its own costs. In particular, intermediary platforms that connect the data gathered from different players could themselves possess market power.[23]

There is a further risk of lock-in to an inefficient standard. To illustrate, assume that a data standard requires all medical data collectors to gather certain types of data at specified intervals, but these intervals are too far apart for the data to be meaningful. While data could be collected at shorter intervals, the standard might send a wrong signal as to the appropriate interval. In addition, data standards can impose high compliance costs on all market players, potentially countering some or all of the competition-driven portability benefits just described.[24] Last but not least, data standards can also negatively affect competition by raising some competitors' costs,[25] and could make coordination and collusion easier.[26]

Data interoperability and portability also raise privacy concerns. The easier it is to share data, the greater the concern that private data will fall into more hands.[27] Portability could also reduce the willingness of potential data subjects to allow their private data to be collected, thereby potentially affecting data collection and innovation.

Data portability can also affect cybersecurity.[28] Integration of databases may enable security systems to more efficiently detect patterns of suspicious activity, and the scale of data may allow algorithms to more rapidly learn from past patterns to detect future attacks.[29] Yet, the more standardized the data, the easier it might be for hackers to access and use it. The potential harm becomes even greater to the extent that

---

20 See, e.g. STUCKE & GRUNES, *supra* note 10, at 279.

21 See, e.g. Eleanor Fox, *We Need Rules to Rein in Big Tech*, CPI ANTITRUST CHRONICLE (Oct. 2020).

22 *Id.* at 166.

23 Michal S. Gal & Niva Elkin-Koren, *Algorithmic Consumers*, 30 HARV. J. L. & TECH. 309, 338 (2017).

24 Peter Swire & Yianni Lagos, *Why the Right to Data Portability likely Reduces Consumer Welfare*, 72 MARYLAND L. REV. 335, 352 (2013); Orla Lynskey, *Aligning Data Protection Rights with Competition Law Remedies? The GDPR Right to Data Portability*, 42 EUR. L. REV. 793, 808 (2017).

25 *Chapter One: Cooperation or Resistance?: The Role of Tech Companies in Government Surveillance*, 13 HARV. L. REV. 1722, 1733-34 (2018) (suggesting that the requirements for data storage applied by the Second Circuit created a comparative advantage to Microsoft relative to its competitors).

26 ARIEL EZRACHI & MAURICE STUCKE, VIRTUAL COMPETITION: THE PROMISES AND PERILS OF THE ALGORITHM-DRIVEN ECONOMY (2016); Michal S. Gal, *Algorithms as Illegal Agreements*, 34 BERKELEY TECH. L. J. 1 (2018).

27 Peter Swire & Yianni Lagos, *Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique*, 72 MARYLAND L. REV. 335 (2013).

28 Security harms do not involve privacy alone but can also engender economic harms, for example through the loss of financial data and identity theft. See, e.g. CLARE SULLIVAN, DIGITAL IDENTITY: AN EMERGENT LEGAL CONCEPT 113–16 (2011).

29 TATIANA TROPINA & CORMAC CALLANAN, SELF- AND CO-REGULATION IN CYBERCRIME, CYBERSECURITY AND NATIONAL SECURITY 14 (2015).

data portability enables the creation of larger, less-dispersed databases, given that the size of the dataset may be positively correlated with the potential harm from security breaches.[30] Finally, an inefficient standard can reduce organizations' ability to detect cyber threats and make its implementation costly.

The costs and benefits of requiring interoperability of data sets and making portability possible are likely to differ among different types of data or its uses. As a result, in some cases it may be better to prevent certain uses of data, including its sharing under certain circumstances. At the same time, in some settings, encouraging portability of data must be accompanied by safeguards – legal, technological or even cultural – that ensure that its overall effects on social welfare are positive.

Should the U.S not take an active role in examining and in some cases possibly even facilitating data standards, American firms might find themselves bound by foreign standards.[31] Given that the European Union has acknowledged the importance of data standards for ensuring a comprehensive data sharing environment,[32] and its market players are currently in the process of setting such standards in order to comply with portability requirements, it is important to ensure that domestic data interoperability and portability considerations are not disregarded.

IV. ANTITRUST ENFORCEMENT AND/OR REGULATION?

Can we rely on the market to create and implement efficient data standards that support interoperability and portability? In a number of settings the answer is in the affirmative, given the large benefits to be had from data standardization. Interestingly, private endeavors have mainly focused on data portability, rather than on data interoperability.[33] Yet, in some settings, significant market failures may prevent socially beneficial data standardization, a vital prerequisite to achieving the benefits of interoperability and portability. Consider the world of music recordings, where songs and other types of music have been saved in a variety of formats (e.g. cartridges, CDs, audio tapes, digital audio recordings, etc.) that can be hard for individuals to use.

This section explores some reasons for this market failure as well as the policy implications. First, the incentives of different market players may differ and may affect their ability to create an efficient standard. Some market participants may favor the status quo with the benefits being high switching costs, greater lock-in and reduced data portability. For some, this characterizes the large platforms – Google, Amazon, Facebook, and Amazon -- incumbents enjoying data-based comparative advantages that cannot be easily matched by others. By preventing the creation of the standard, the claim is that incumbents essentially raise their rivals' costs relative to their own.

Second, even if a standard is voluntarily created, its content may serve the interests of some market players and not others. Concerns arise from the private interests of those involved in setting the standard, especially given the knowledge that competitive entry may involve substantial sunk costs. Furthermore, the chosen standard may impose costs as well as benefits in the rivals of its creators.[34]

Third, collective action problems might lead market players not to make portability possible, even when it is beneficial for all of them to do so. In the absence of an arbiter, the market may be sufficiently fragmented that no single approach gains critical support, leading to a patchwork of inconsistent data standards that slow data flows.[35] Furthermore, there might be insufficient time for deliberation before the market sets on its course. Most importantly, the uncertainty resulting from the fact that users cannot be assured that others will follow their move to the new standard, creates a coordination problem.[36] Coordination incentives could also be limited by lack of knowledge among data collectors about the data's potential uses and concerns about the obstacles to integrating it with other types of data. Antitrust concerns, too, could limit incentives to standardize. And, the creation of efficient data standards might be inhibited by internal constraints, short-term strategic conduct, or historical legacies.

---

30 Wolfgang Kerber & Heike Schweitzer, *Interoperability in the Digital Economy*, 8 J. Intell. Prop. Info. Tech. & Electronic Comm. 39, 54 (2017).

31 See generally Anu Bradford, *The Brussels Effect*, 107 Nw U. L. Rev. 2 (2012).

32 Article 29 Data Protection Working Party, Guidelines on the Right to "Data Portability" 2 (2017), http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233.

33 See, e.g. Lynskey, at 793; Data Portability Project, http://www.dataportability.org (last visited Nov. 3, 2020); & Open Data Institute, https://theodi.org/ (last visited Nov. 3, 2020).

34 Stanley M. Besen & Joseph Farrell, *Choosing How to Compete: Strategies and Tactics in Standardization*, 8 J. Econ. Persp. 117, 128 (1994).

35 Kevin Werbach, *Higher Standards: Regulation in the Network Age*, 23 Harv. J. L. & Tech. 179, 201 (2009).

36 See Joseph Farrell & Garth Saloner, *Coordination Through Committees and Markets*, 19 Rand J. Econ. 235, 236 (1988).

Even if the portability of data that serves the interests of all market players is achieved, private standard-setters may disregard the positive spillovers they create on data subjects, on firms in other markets, and on social welfare. An inherent tension also exists between temporal beneficiaries of data analysis: while tomorrow's users may benefit from past data collection, their gains are not always easily shared with the collectors of such data.

Market failures may also arise with regard to the implementation of an acceptable standard. There is arguably an important regulatory role in the acknowledgment, evaluation, and – in the right cases – the possible facilitation of data portability. The potential benefits from increased uses of data, as well as the costs accruing from the potential loss of international competitiveness and from the continuing use of a patchwork of (inefficient) standards, should act as a catalyst for data portability issues to be seriously considered.

As an initial first-stage effort, Congress and the competition authorities should carefully study market dynamics and characteristics to identify where data portability's benefits outweigh its costs. Such costs include the costs of standard setting, implementation, and oversight, of compliance with the standard, and lock-in to an inefficient standard. The need for study is strengthened by the fact that the current situation is characterized by a patchwork of inconsistent legacy data collection and organizational methods, developed over time by various market players, which are not particularly conducive to data integration.

The competition authorities are well positioned to analyze the pros and cons of data portability. They have or can acquire the appropriate technical expertise. They have the ability to understand the implications of their decisions on all market players, to evaluate whether industry standards are economically efficient, and to assess whether the market could and would develop timely and efficient standards without governmental intervention. However, there is a case to be made for the creation of a regulatory overlay, perhaps through the promulgation of a set of regulatory constraints under which the competition authorities should operate.

Creating an ecosystem of standards that can work in different contexts, and that can interoperate where required, is likely to also require consultation with industry, or even a coordinated governance process that includes the participation of market players. Both suggestions build on the fact that market players often have substantial knowledge and understanding of existing technical needs and the merits of a variety of possible solutions. The particular governmental agency that takes the lead in doing so, and the specific agenda it pursues, may vary across industries.

Once it is established that allowing for data portability will likely increase social welfare, it is important to facilitate the creation of efficient data standards. Regulators face a range of options with regard to how standards can be set, each with its own costs and benefits. These include adopting private solutions, establishing standard-setting organizations (SSOs) or determining standards themselves. The preferred regulatory model may differ among industries and among types of data, depending on the relative competence of different standard setters, the extent of divergence between private and social interests, and the way such a divergence might affect the costs of portability. Yet it seems that in most cases a supervised delegation to an industry-based SSO, comprised of professional data scientists, will be more advantageous than performing the task by a new governmental entity. While regulators play an important role in determining when market failures prevent the creation of welfare-enhancing data standards, they generally have less competence in evaluating the standards that will work best in a given market setting. Where private SSOs are preferred, the regulator may need to set and enforce some basic rules for their operation.

Once a data standard is agreed upon, the regulator must decide how to facilitate its adoption. Options include setting best-practices, mandating the adoption of data standards, and creating soft incentives for their adoption.[37] It might come as no surprise that the Data Transfer Project undertaken in June 2018 by Microsoft, Google, Facebook, and Twitter, which sets a standard to enable user-initiated data portability among project participants, was initiated amidst increased calls for the government to reign in the power of large digital firms resulting from the control of data.[38]

It is noteworthy that in some situations the government may have no choice but to set data standards if it is to make portability work. This might be the case where the government collects and organizes data internally (such as meteorological, demographic or legal data), or where it contracts with others to provide it with certain types of data.

---

37 The Office of the National Coordinator for Health Information Technology, for example, releases an annual list of best available standards, to be used by technology developers and to inform coordinated governance efforts. See Office of the Nat'l Coordinator for health info. Tech., 2015 Interoperability Standards Advisory 1 (2015).

38 See also Greg Fair, *Our Work to Move Data Portability Forward*, The Keyword (Sept. 21, 2020), https://blog.google/technology/safety-security/data-portability (discussing Googles efforts to improve data portability through the Data Transfer Project). Of further note, Google Takeout was created on June 28, 2011 to allow users to export their data from most of Google's services. See www.wikipedia.com (las visited Nov. 8, 2020).

V. CONCLUSION

There are substantial benefits, along with some potentially significant costs, to increasing data portability. The private sector has been active in making efforts, individually or jointly, to improve data portability. Nevertheless, private benefits and social benefits are not fully aligned and there is a clear role for intervention by the public sector. Adding a regulatory overlay to our current regulatory and enforcement authorities that recognizes the potential effects of data portability is appealing. The value of adding such an overlay substantially, short of the creation of an entirely new governmental entity, makes sense. Of course, given the costs and risks involved in intervening in the market, caution is required before any such intervention.