ACCESS BARRIERS TO BIG DATA

Daniel L. Rubinfeld and Michal S. Gal[**]

An updated version to be published in the Arizona Law Review (2017)

INTRODUCTION

'Big data' was recently declared to represent a new and significant class of economic asset that fuels the information economy.[1] While data were always valuable in a range of economic activities, the advent of new and improved technologies for the collection, storage, mining, synthesizing, and analysis of data has led to the ability to utilize vast volumes of data in real-time in order to learn new information. Big data is a game-changer since it allows for regularized customization of decision-making, thereby reducing risk and improving performance. It also changes corporate ecosystems by moving data analytics into core operational and production

[**] Robert L. Bridges Professor of Law and Professor of Economics Emeritus, U.C. Berkeley and Professor of Law, NYU; and Professor and Director of the Forum for Law and Markets, University of Haifa Faculty of Law. The authors thank Ron Bekkerman, Avigdor Gal, Daniel Sagis, Kira Radzinsky, and Sagie Rabinovitz for helpful discussions on big data issues in the real world and Konstantina Bania, Wolfgang Kerber, Warren Grimes, Danny Sokol, Nicolas Petit and Hal Varian as well as participants in the ASCOLA conference and the Cyber Regulation Forum at the University of Haifa for most helpful comments on previous drafts. The authors would also like to thank Orel Asulin, Lior Frank, Meital Sadovnick and Ido Sharipra for excellent research assistance. Any mistakes or omissions remain the authors'. Many thanks to the Minerva Center for the Rule of Law Under Extreme Conditions- Cyber Regulation for funding.

[1] WORLD ECONOMIC FORUM, BIG DATA, BIG IMPACT: NEW POSSIBILITIES FOR INTERNATIONAL DEVELOPMENT (2012); Kenneth Cukier, *Data, Data Everywhere*, THE ECONOMIST, February 27, 2010 (consumer data are becoming "the new raw material of business: an economic input almost on a par with capital and labor").

functions, and it enables the introduction of new products (e.g., self-driving cars and digital personal assistants such as Siri).[2]

The data gathered are wide-scoped and varied, ranging from natural conditions to Internet user's preferences. It can be gathered using digital as well as non-digital tools. The Internet, including the internet-of-things which places trillions of sensors in machines all around the world, generates vast amounts of data.[3]The "Data Religion" practiced by many, which worships sharing and transparency and whose main principle is "more transparent data is better,"[4]coupled with the willingness of many users to share data for only a small benefit,[5] further increases the amount of data that can be collected. This, in turn, has led to the creation of entities which specialize in the collection, management analysis, and/or visualization of big data and which use the data themselves or broker it. These entities compete for users' attention,[6]since the wealth of information creates a "poverty of attention."[7] Indeed, a whole range of products and services is provided free of direct charge, in exchange for the ability to harvest user's data.[8]

Yet data, by itself, are often of low value. What gives data value is its analysis, turning unstructured bits and bytes into information and derived information (i.e., applying reasoning mechanisms to create new information that cannot be gathered directly from the data), in order to turn it into actionable information, both descriptive as well as predictive. Rapidly advancing techniques of data science such as natural-language

[2] *See, e.g.,* Erik Brynjolfsson, Lorin M. Hitt, and Heekyung Hellen Kim, *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?* (April 2011), *available at* http://ssrn.com/abstract=1819486 (effective use of data and analytics correlated with a 5- 6% improvement in productivity, as well as higher profitability and market value); Thomas H. Davenport, Paul Barth and Randy Bean, *How 'Big Data' is Different*, MIT SLOAN MANAGEMENT REVIEW (Fall 2012).

[3] *See, e.g.,* OECD, SUPPORTING INVESTMENT IN KNOWLEDGE CAPITAL, GROWTH AND INNOVATION (2013).

[4] *See, e.g.,*Niva Elkin-Koren, *An Intimate Look at the Rise of Data Totalitarianism*, JOTWELL (April 21, 2015); YUVAL NOAH HARRARI, THE HISTORY OF TOMORROW (2016).

[5]*See, e.g.,* Alessandro Acquisti, Leslie K. John & George Loewenstein, *What Is Privacy Worth?* 42(2) J. OF LEGAL STUD. 249 (2013).

[6]*See, e.g.,* David S. Evans, *Attention Rivalry Among On-line Platforms and its Implications for Antitrust Analysis*, 9(2) J. OF COMP. L. AND ECON. 313 (2013).

[7]*See, e.g.,* Herbert A. Simon, *Designing Organizations for an Information-rich World, in* COMPUTERS, COMMUNICATIONS, AND THE PUBLIC INTEREST (M. Greenberger ed., 1971) 38; CARL SHAPIRO AND HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY (2nd ed., 2013).

[8]*See, e.g.,* Michal S. Gal and Daniel L. Rubinfeld, *The Hidden Costs of Free Goods: Implications for Antitrust Enforcement*, 80 ANTITRUST L. J. 401 (2016).

processing, pattern recognition, and machine learning, together with traditional tools such as statistics, are utilized in order to mine valuable information from data.[9] This, in turn, creates a virtuous circle between the incentives to collect new data and advances in its synthesis and analysis.

Such data-based information has provided an important input into decision-making in our modern economy. It is valuable to commercial firms (e.g., enabling them to create better products or services, to better target consumers by creating tailor-made ads to products they are likely to buy, to price discriminate among consumers based in their revealed preferences, and to have a deeper, richer advance knowledge of potential employees and suppliers),[10] to governments (e.g., locating terrorist cells, predicting and possibly reducing the harmful effects of disease outbreaks, climate impacts, economic development, job losses, and even governmental corruption), and to individuals (e.g., enjoying better products, or knowing what others think of certain issues).

In such a world, access to data and to the information based on it become strategic and valuable assets.[11]Those who enjoy more portholes from which to gather data, who have a substantial database to which they can compare new data, or who possess unique data synthesis and analysis tools, may enjoy a competitive comparative advantage. The value of such information is apparent from the number of acquisitions of data sources and the prices paid for them,[12] the vast investment in data gathering, and in the creation of better machine learning tools. Further indications of its value include the reluctance of data-collecting firms to provide users with the option of preventing the collection of personal data as well as their reluctance to

[9] *See, e.g.,* Hal R. Varian, *Big Data: New Tricks for Econometrics*, 28(2) JOURNAL OF ECONOMIC PERSPECTIVES 3 (2014).

[10] *See, e.g.,* Brynjolfsson, Hitt & Kim, *supra* note 2; Michael A Salinger and Robert J. Levinson, *Economics and the FTC's Google Investigation*, 46 REVIEW OF INDUSTRIAL ORGANIZATION 25 (2015); Alessandro Acquisti and Hal R. Varian. *Conditioning Prices On Purchase History*, 24(3) MARKETING SCIENCE 367 (2015).

[11] *See, e.g.,* OECD, *supra* note 3, at 322; JAMES MANYIKA., MCKINSEY GLOBAL INST.,BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY 13 (2011); OECD, DATA-DRIVEN INNOVATION FOR GROWTH AND WELL-BEING: INTERIM SYNTHESIS REPORT10 (2014).

[12] *See, e.g.,* Letter From Jessica L. Rich, Director of the Federal Trade Commission Bureau of Consumer Protection, to Erin Egan, Chief Privacy Officer, Facebook, and to Anne Hoge, General Counsel, WhatsApp Inc. (April 10, 2014), https://www.ftc.gov/public-statements/2014/04/letter-jessica-l-rich-director-federal-trade-commission-bureau-consumer; Data-Driven Innovation, *ibid*, at 94 (in sectors related to data, "the number of mergers and acquisitions (M&A) has increased rapidly from 55 deals in 2008 to almost 164 deals in 2012.").

enable data portability, and the competitive and litigation-related wars relating to data access. Indeed, data may create comparative advantages not only for firms and individuals, but also for countries.[13]

Given the central role that big data plays in our modern knowledge-based economy, an analysis of the way markets for big data operate and how they affect welfare is essential. Yet, while numerous articles and reports are being written on public policy issues,[14] no real attempt has been made to analyze in depth the economic characteristics of the markets for big data. Rather, most of the analysis is based on broad, unverified assumptions about how such markets operate. To illustrate, Tucker and Wellford argue that big data does not create a significant barrier to entry. They base their claim, inter alia, on the non-exclusive and non-rivalrous nature of data and a claimed ease of collecting it, while disregarding many potential entry barriers.[15]Other scholars argue that the harm created by big data focuses mainly on harm to privacy.[16]Yet these conclusions are based on the limited existing economic studies on big data which often focus on one specific market (most commonly on search engines or on personal data markets).[17]the characteristics of which do not necessarily carry over to

[13]Brad Brown, Michael Chui, and James Manyika, *Are You Ready for the Era of "Big Data?* (October 2011), http://www.mckinsey.com/insights/strategy/are_you_ready_for_the_era_of_big_data.

[14]*See, e.g.,* EUROPEAN DATA PROTECTION SUPERVISOR, PRIVACY AND COMPETITIVENESS IN THE AGE OF BIG DATA: THE INTERPLAY BETWEEN DATA PROTECTION, COMPETITION LAW AND CONSUMER PROTECTION IN THE DIGITAL ECONOMY (2014); OECD Report, *supra* note 11; FRENCH AND GERMAN COMPETITION AUTHORITIES, COMPETITION LAW AND DATA (2016) http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2. (enumerating ways in which big data can potentially harm competition); UK COMPETITION AND MARKETS AUTHORITY, THE COMMERCIAL USE OF CONSUMER DATA (2015), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/435817/The_commercial_use_of_consumer_data.pdf .

[15] Darren S. Tucker and Hill B. Wellford, *Big Mistakes Regarding Big Data* 14 ANTITRUST SOURCE 6, 6 (2014).

[16] *See also* Maureen K. Ohlhausen, and Alexander P. Okuliar, *Competition, Consumer Protection, And the Right [Approach] To Privacy,* 80 ANTITRUST L. J. 121; James C. Cooper, *Privacy and Antitrust: Underpants Gnomes, the First Amendment, and Subjectivity* 20 GEORGE MASON L. REV. 1129 (2013).

[17]*See, e.g.,* Andres V. Lerner, *The Role of 'Big Data' in Online Platform Competition,* (2014), http://ssrn.com/abstract=2482780.

other big-data markets, and on the assumption that big data markets exhibit low entry barriers because of the inherent characteristics of big data.[18]

Other scholars, including Stucke and Grunes, Shelanski, and Geradin and Kuschewsky, challenge these conclusions.[19] While the studies are thoughtful, none engage or build upon in-depth economic studies of the characteristics of big data markets. At the end of the spectrum, a recent OECD report concluded that markets which are based on big data are likely to be concentrated, and with scenarios of "winner takes it all," but did not base its conclusions on in-depth economic analysis.[20]Indeed, Sokol and Comerford argue that "the scholarly case for [a theory of harm to competition] has not yet been adequately established."[21] In their recent book, Stucke and Grunes advance the literature by focusing mainly on the network effects of big data, which leads them to conclude that "[t]he economics of data ....favor market concentration and dominance."[22] While important, network effects are not present in the same degree in all big data markets. Moreover Stucke and Grunes focus mainly on data collection.

Finally, the literature often comments on "big data markets" as a whole, without asking whether there are appropriate sub-groups with distinctive and policy-relevant characteristics. Similar shortcomings characterize reports issued by some antitrust authorities.[23]

Given the importance of big data to our economy, and the broad range of normative issues that it raises, there is a need to improve upon the current state of knowledge. This need is strengthened by the Collingridge dilemma, according to which by the time a technology's undesirable consequences

[18] D. Daniel Sokol & Roisin Comerford, *Does Antitrust Have a Role to Play in Big Data?*, *in* CAMBRIDGE HANDBOOK OF ANTITRUST, INTELLECTUAL PROPERTY AND HIGH TECH (forthcoming).

[19]Maurice E. Stucke & Allen P. Grunes, *Debunking the Myths over Big Data and Antitrust*, CPI ANTITRUST CHRONICLE (May 2015); Maurice E. Stucke and Allen P. Grunes, *No Mistake About It: The Important Role of Antitrust in the Era of Big Data*, ANTITRUST SOURCE (2015); Howard A. Shelanski, *Information, Innovation, and Competition Policy for the Internet*, 161 U. PA. L. REV. 1663, 1679 (2013) ("[C]ustomer data can be a strategic asset that allows a platform to maintain a lead over rivals and to limit entry into its market."); Damien Geradin & Monika Kuschewsky, *Competition Law and Personal Data: Preliminary Thoughts on a Complex Issue* (2013) http://ssrn.com/abstract=2216088; Frank Pasquale, *Paradoxes of Digital Antitrust: Why the FTC Failed to Explain Its Inaction on Search Bias*, HARV. J. J. & TECH. (Occasional Paper Series)(July 2013).

[20]OECD, DATA-DRIVEN INNOVATION, *supra* note 11, at 7.

[21] Sokol and Comerford, *supra* note 18.

[22]MAURICE E. STUCKE AND ALLEN P. GRUNES, BIG DATA AND COMPETITION POLICY (2016), p. 337 (hereinafter: "BIG DATA").

[23] THE COMMERCIAL USE OF CONSUMER DATA, *supra* note 14.

are discovered, the technology is often so much part of the economic and social fabric, that its control is extremely difficult.[24]

Accordingly, this article seeks to take a first step towards filling this void. Part I explores the four primary characteristics of big data: volume, velocity, variety, and veracity and their effects of the value of data. Part II analyzes the different types of barriers that limit entry into the different links of the data value chain. These barriers determine, to a large degree, the effect on competition and welfare. They also serve as a crucial step in delineating markets for a competitive analysis. While some of the barriers explored are well known, others are not. We accentuate differences among big data markets. Accounting for those differences, we suggest a set of smaller relevant markets, in accordance with their specific entry barriers: technological, legal, and behavioral. Among the questions to be addressed are whether data collection creates entry barriers with strong first-mover advantages or bottlenecks; the extent of returns to scale and scope in data collection and analysis, including user-quality and monetization-quality network effects; how attention rivalry, two-sided markets, multi-homing, and the data's non-rivalrous nature may affect some markets; and whether control over some types of data is essential to determine whether there could be meaningful competition. The analysis is based on the existing literature on technological and legal barriers, as well as on several in-depth interviews of players in big data markets that we performed.

In Part III, we tie together the characteristics of big data markets including potential entry barriers, to analyze their competitive effects. The analysis centers on those instances in which the unique characteristics of big data markets lead to variants in the more traditional competitive analysis. Part IV concludes.

Our conclusions have important implications for public policy. Our analysis suggests that the unique characteristics of big data have an important role to play in analyzing competition and in evaluating social welfare. The issues could hardly be timelier, as questions regarding entry to big data markets and access to data are already taking center stage in our information-driven economy.

PART I: CHARACTERISTICS OF BIG DATA

'Big Data' is a generic name for data that share several characteristics with regard to its aggregation, rather than content. Its main characteristic,

[24] DAVID COLLINGRIDGE, THE SOCIAL CONTROL OF TECHNOLOGY (1980).

as its name signifies, is its volume. A simple definition relates to amounts of data that cannot be analyzed by traditional methods; rather, it requires the establishment of a unique platform that can manage substantial volumes of information in a reasonable timeframe.[25]The concept of 'big data' is thus a moving target, given that the volume of data captured under the definition today may change over time as storage and analysis capabilities grow. What is possible today with terra-bytes was not possible, only a few years ago, with gigabytes.

The fact that the term "big data" relates to volume rather than content leads to several observations. First, big data can relate to a wide range of content (e.g., earthquakes, calories, consumer preferences, etc.), and consequently may be used as inputs in a variety of markets. Second, there is no market that requires generic big data, as such. Rather, different markets often need different types of big data as inputs. For example, both a book publisher and a sports-car manufacturer would find information on the preferences of potential buyers (e.g., income and spending habits) to be valuable. Yet while the book publisher is less interested in income, given the relatively low prices of books, a sports car manufacturer has a greater interest in it. Therefore, the users of big data often do not compete with each other, depending on the product and geographic market in which they operate. This also implies that simply because vast amounts of data are collected, one cannot conclude that data are fungible.[26]

This leads to a third observation, that due to its non-rivalrous nature, the same dataset can be useful to a variety of users and consequently is likely to have different value for different users.[27] Fourth, the collectors, aggregators and analyzers of big data do not necessarily compete, if they adopt or analyze different types of data (e.g., one collects data on geological phenomena and another on drug use). Fifth, the data need not be collected from the same sources or in the same way in order to be competitive (e.g., collection from the Internet, from wearables, from smart appliances, or from personal interviews). Finally, at least some of the data collected is a by-product of other productive activities, a fact which does

[25]*See, e.g.,* Manyika et al., *supra* note 11, at 1 ("'Big data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.").

[26] Se also BIG DATA, *supra* note 22, p. 79.

[27] See also FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

not affect the importance of the data, but which could affect the ability of multiple firms to collect it. Taken together, these facts imply that determining whether a big data collector possesses market power, mandates that one define the market in which the data collector operates, as well as the use(s) of such data, much like any other market analysis.

While big data markets may differ substantially from one another, most big datasets share four main characteristics, which contribute to their value: Volume, Velocity, Variety, and Veracity.[28]*Volume* is arguably the most important characteristic. Developments in collection, storage, and analytical capabilities have exponentially increased the volume of data that can be collected and analyzed. *Velocity*- sometimes referred to as the "freshness" of the data -relates to the speed of change. Especially in dynamic markets, new data might render some old data obsolete.

Variety is characterized by the number of different sources from which the data are gathered. Indeed, the vast amount of data often implies that several types of data are gathered together, for example, information about consumers' location, preferences, circle of friends, etc. Alternatively, data sources can be multi-varied. Such sources can be human actions (e.g., a user of Facebook or Whatsapp) or machines (internet of things, wearables). Sources can also be primary (a post on a blog) or secondary (data already gathered by someone else).The integration of data from different sources may significantly increase the value of the dataset. Variety can also relate to the time period covered by the data. Finally, *veracity* relates to the truthfulness of the data –in essence, its accuracy. This characteristic can relate to the accuracy of the building blocks of the database (implying that not all data are equal), or to the database as a whole, since what we lose in accuracy at the micro level might be gained in insights at the macro level.

The importance of each of these characteristics might differ among the myriad of markets in which big data serves as an input. For example, where velocity is of small importance relative to the other three parameters, the data might not have to be constantly updated. Rather, old data can serve as a sufficiently effective input for firms competing in the market.

These four characteristics relate to the collection of big data and serve as the basis for its value. Yet what often gives big data its real advantage is not these characteristics alone, but rather the ability to synthesize and

[28] Mark Lycett, *Datafication: Making Sense of (Big) Data in a Complex World*, 22 EUROPEAN JOURNAL OF INFORMATION SYSTEMS 381 (2013); OECD, SUPPORTING INVESTMENT IN KNOWLEDGE CAPITAL, GROWTH AND INNOVATION (2013); EXECUTIVE OFFICE OF THE PRESIDENT, PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE (2014).

analyze the big data, characterized by the four V's, in ways not previously available, thereby creating meta-data (sometimes known as information-on-information).Advancements in data science have contributed to the ability to learn fast and deep from big data by analyzing correlations among variables. Such advancements include data mining techniques such as association analysis, data segmentation and clustering, classification and regression analysis, anomaly detection, and predictive modeling. They also include Business Performance Management (BPM), which helps one to analyze and visualize a variety of performance metrics based on the data.[29]These technological advancements allow algorithms to extremely quickly and efficiently find patterns and make predictions in ways that no human analyst could perform (without getting sick, tired, or bored!).Furthermore, the change in quantity has led to a change in quality: due to its size, big data provides insights which are not possible from traditional data. Finally, new analytical tools enable algorithms to automatically "learn" from past actions, thereby creating a positive feedback loop, in order to become more efficient in mining the data.

Big data thus leads to more informed decision-making which, in turn, might lead to more efficient decisions by firms, individuals, and governments. Accordingly, the term "big data" is often used as a proxy not for the rough data gathered and collected, but rather for the meta-data which flow from the synthesis and analysis of initial rough datasets. Such data are central to operations in many markets. As the OECD observed, "Big data now represents a core economic asset that can create significant competitive advantage for firms and drive innovation and growth."[30]

Big data also suffers from some limitations. While technological ones will be explored in the next part, here we emphasize one inherent limitation: its analysis offers correlations, but does not allow one to determine causality.[31]This, in turn, creates an increased risk of "false discoveries"[32]which, in turn, might affect the quality of the information and also imposes costs on those wrongly categorized. We note, however, that data scientists are in the process of devising algorithms which make more accurate predictions and verify them.

[29] Chen et al, *Introduction: Business Intelligence Research*, 36 MIS QUARTERLY 1165 (2012).

[30] OECD, SUPPORTING INVESTMENT IN KNOWLEDGE CAPITAL, GROWTH AND INNOVATION, 10 October 2013, p. 319. *See also* additional studies cited in BIG DATA, supra note 22, ch.4.

[31] VIKTOR MAYOR-SCHONBERGER AND KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK (2013).

[32] *See, e.g.,* FTC, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION (2016).

Other concerns also arise, and these can be expected to grow in significance over time. Individuals' conduct may change once they are aware of how the data are gathered and used. For example, if different prices are set for first-time users and repeat users, then users might erase their search histories in order to enjoy first-user benefits. But more importantly, big data might increase harms to privacy and increase disparities or biases by categorizing certain (groups of) consumers in ways that can result in exclusion or discrimination. For example, firms could use big data to exclude low-income and underserved communities from certain beneficial credit and employment opportunities.[33] At the same time, however, big data can open up opportunities for weak groups in society.[34]Alternatively, it might be used for individualization of products and prices, which raises social welfare issues.[35] Another potential harm involves changes in consumer conduct that are influenced by data-based predictions which, in turn, might harm consumer welfare. This is a subject for another paper.

It is interesting to note that because of the value big data can generate, firms sometimes offer high quality or subsidized products or services to consumers, in exchange for the harvesting and use of data.[36] This, in turn, may increase access of more groups in society to helpful products or services which they could not have otherwise afforded. Yet the overall welfare effect of this exchange depends on the positive as well as the potential negative effects of the use of the collected data, as elaborated above.

Big data markets are comprised of several links along the data value chain as depicted in Diagram 1: Collection – Storage –Synthesis & Analysis - Decision-making.[37]

[33]*See, e.g., ibid.*, at 8-12; Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV.(April 1, 2013); Cynthia Dwork & Deirdre Mulligan, *It's Not Privacy and It's Not Fair*, 66 STAN. L. REV. ONLINE 35,36–37 (2013).

[34]One example involves RiskView, a credit score firm which is based on data from non-traditional indicators, thereby allowing consumers who may not have access to traditional credit, to be given better access to financial markets. FTC REPORT, *supra* note 32, at 6, 27 ("research suggests that big data offers both new potential discriminatory harms and new potential solutions to discriminatory harms.").

[35] For an analysis of costs and benefits of big data, see also GERMAN MONOPOLIES COMMISSION, COMPETITION POLICY: THE CHALLENGE OF DIGITAL MARKETS, Special Report No 68 (2015).

[36]David S. Evans and Richard Schmalensee, *Industrial Organization of Markets with Two-Sided Platforms*, 3 COMPETITION POLICY INTERNATIONAL 154 (2007).

[37]See also FTC REPORT, *supra* note 32, at 3.

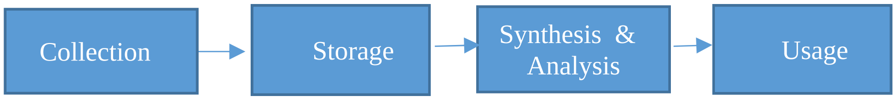

Diagram 1: The data value chain

Collection relates to the extraction of the data; Storage relates to the tools for transformation, load and storage of data, which are essential for organizing data in a database; Synthesis and Analysis relate to the integration of different types of data and to the analytical processing of the data in order to find correlations. The last link involves usage of the data-based information for decision-making in relevant markets.

The data-value chain implies that when talking about big data markets, we must be clear as to the link(s) of the data value chain that are being analyzed. This seemingly simple observation is not always followed, especially in literature relating to regulatory policy. In this article we shall refer to each link separately, unless we use the term "big data markets," which will relate to all links of the chain.

PART II: ACCESS BARRIERS INTO BIG DATA MARKETS

A. OVERVIEW

The characteristics of big data noted above serve as a basis for the next step in our analysis: exploring entry barriers to big data markets. Entry barriers are defined as barriers to the entry or the expansion of firms in relevant markets. Some barriers are unique to big data (e.g., storage of large volumes of data), and some are also relevant to traditional data. The specific characteristics of the data needed for each market in which it serves as an input affect the height and type of entry barriers.

Traditionally, entry barriers are seen as coming from the supply side, driven for example by the presence of substantial economies of scale that have been achieved by one or more incumbent firms, but which are not easily achieved by a new entrant. However, barriers can also arise from the demand side. We offer commentary with respect to both sources of barriers.

Entry barriers to big data markets can be analyzed in accordance to the relevance of the restraint to each of the four characteristics of big data, the type of the restraint (technological, legal, behavioral, etc.), or the stage in the data value chain (collection, storage, synthesis and analysis, usage). We have chosen to organize our analysis in accordance with the latter, while referring to the other categorizations, where relevant. This enables us to emphasize the fact that access to big data requires more than its collection.

Most of the barriers analyzed apply to the entry or expansion of firms in each part of the data value chain. Some also apply to data portability. Portability is an important way to access data that someone else collected, stored, or analyzed. It is possible due to the data's non-rivalrous nature, "a key factor for effective competition" in data markets.[38]Note that in this part of the analysis we identify entry barriers, regardless of their potential offsetting welfare-enhancing goals or whether they can be overcome by regulatory means.

B. BARRIERS TO THE COLLECTION OF BIG DATA

This section focuses on the first stage in the data value chain- the collection of the data. We identify three types of barriers: technological, legal, and behavioral, which can exist in parallel and can reinforce one another. This might happen, for example, where the use of an efficient technological route is prohibited by law, and (some) market players must find an alternative, less-efficient route.

1. Technological Barriers

a. Uniqueness of the data or the gateways to it

It is often argued that due to its non-rivalrous nature and the fact that sources of data are often abundant, big data can be easily and inexpensively collected by many firms in parallel. This is true with regard to publicly available data which is freely available to anyone (e.g., the martindale.com attorney information list), or a user's location which can be simultaneously observed by many smart-phone applications at a very low cost. Furthermore, when the data are useful as an aggregate, the data can be often collected from different sources, thereby enabling easy substitution between alternative data sources. For example, data used to determine the average speed of drivers in on dark streets, can be collected from a variety of locations. Where the cost of collecting alternative data is not prohibitive, there is no barrier. This was recognized, for example, in the FTC's investigation of the Google/DoubleClick merger. The FTC concluded that the firms' user data was not a barrier to entry for online advertising, since "neither the data available to Google, nor the data available to DoubleClick, constitutes an essential input to a successful online advertising product."[39]

[38] EU COMMISSION, IMPACT ASSESSMENT REPORT ON THE PROPOSED GENERAL DATA PROTECTION REGULATION (2012).

[39] Statement of Federal Trade Commission 12, Google/DoubleClick, FTC File No. 071-0170 (Dec. 20, 2007). See also EU Commission decision Google/Doubleclick, Case COMP/M. 4731 [2008] OJ C 184/6.

Nonetheless, unique access points to unique data may lead to situations in which the data cannot be easily replicated. This might be the case if the data were created as a result of a distinctive interaction. Take, for example, Facebook's analysis of how emotions, expressed and created by the information posted on its site, affect people's conduct. Duplicating such emotional interactions may be difficult and costly.[40] The creation of such vantage collection points may be the primary goal of the data collector, or it might be a side benefit from an otherwise-focused activity. As elaborated below, two-level entry problems, where such unique access points are part of a unique activity, may further increase barriers.

Another barrier may be temporal, relating to the point in time in which the competitor started gathering data. To illustrate, a collection of aerial maps before a natural disaster cannot be replicated once the disaster occurred.[41] Temporal advantages can also arise when one collector has unique knowledge at an important point in time. Search engines illustrate this point: they can easily identify consumers engaged in active search of a certain product.[42]

Finally, unique gateways for data collection might also create technological entry barriers. For example, in some third-world countries in which computers are not commonplace, cell phones are the main device for the use of the Internet, thereby creating a technological advantage to cell phone operators in the collection of data. A more subtle gateway barrier involves preinstalled applications that also gather data, which due to the default option and users' status-quo bias, make it harder for other applications to replace them.[43] To give an example, when a data-collecting service application is preinstalled on major platform (e.g., Android), this may create an entry barrier for other data collectors that provide similar services.

[40] *See, e.g.,* Dana Boyd and Kate Crawford, *Six Provocations for Big Data* (2011), http://ssrn.com/abstract=1926431

[41] This example is partially based on the proposed acquisition of Vrisk by EagleView. EagleView Technology had a 90 percent share of the aerial photographic business with respect to the evaluation of damaged homes in tort insurance cases. In part on that basis, the FTC sued to block the acquisition of a small entrant, Verisk. *See, e.g.* ,Deborah Feinstein, *Big Data in a Competition Environment*, 5 COMPETITION POLICY INTERNATIONAL (2015).

[42] FRENCH AND GERMAN REPORT, *supra* note 18, 44.

[43] BIG DATA, *supra* note 22, p. 96.

b. Economies of Scale, Scope and Speed

Technological supply side barriers can arise if incumbent firms have achieved substantial economies of scale or scope through investments which are partially or wholly sunk. A firm that is unable to achieve minimum viable scale (the scale necessary to obtain a competitive rate of return) will find it profitable to find an alternative investment. In addition, where substantial "learning by doing" economies exist, the need to make sunk expenditures may make entry prohibitive. In either of these instances, economies have the potential to deter entry if the data extracted are not redeployable in other markets.

Scale and scope economies in data collection can arise from a variety of potentially cumulative factors. They may flow from the fixed costs of creating devices for data collection (e.g., antennas, crawlers, interview facilities, cookies, etc.), monitoring, and data extraction. Indeed, often the cost of putting in place an infrastructure for data collection and analysis may generate high fixed costs, while the costs of data extraction may be low. Scope economies may also arise from synergies in data analysis.[44] Google's acquisition of Nest Labs, a manufacturer of interactive thermostats which use sensors to train themselves to a user's schedule, serves as an example. The data sent by the thermostat and other home smart devices helps Google create a fuller picture of users' conduct, stretching the power of Google's algorithms beyond the web and into the internet of things.[45] Where scope economies are substantial, they might create entry barriers to those that have access to only one data source.

Basic statistics provides support for the likely presence of economies of scale in data collection. If the data are drawn randomly from almost any underlying distribution, the Central Limit Theorem tells us that the accuracy (given by the standard error) of estimates of the mean of the distribution will increase with the square root of the number of items in the database.[46] Qualitatively similar results apply to other distribution statistics (median, various quantiles).[47]

[44] One example involves Google Flu Trends, which applies data analytics tools to in order to generate meta data which estimates the spread of the virus. Google updated its algorithm in 2014 to incorporate not only Web-based search results, but also data from the Centers for Disease Control and Prevention in the U.S., to boost its algorithm's accuracy. Bloomberg News, *Google Undates Flu Trends to Improve Accuracy* November 1, 2014 http://www.bloomberg.com/news/articles/2014-10-31/google-updates-flu-trends-to-improve-accuracy.

[45] FTC, Google Inc; Nest Labs Inc., Early termination notice 20140457, 4 February 2014.

[46] The same result applies to the magnitude of any irreducible error not account for in the estimation of the mean. For details, see the discussion of the bias-variance tradeoff at https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff.

[47] The accuracy point is even more general. It holds as long as the underlying distribution has finite moments. Furthermore, evidence from studies of machine learning shows substantial

In addition to traditional scale and scope economies- which relate to the data's volume and variety, the velocity of data collection might create what can be called "economies of speed." [48] "Nowcasting" provides an illustrative example.[49]Nowcasting is the capacity of a company to use the velocity at which a dataset grows to discern trends well before others. Nowcasting enables firms not only to track trends in users' conduct in real-time, but also trends in (potential) competitors' conduct, and to respond more quickly.[50]

Due to big data's multivariate nature, the extent of economies is affected by all the four technical characteristics of big data, alone and in combination. To illustrate, Google is able to predict automobile sales, unemployment claims, and travel destination planning thanks to the number and type of queries entered by its users in a particular geographic area, thereby enjoying a temporal advantage. In other cases the data's four characteristics might cancel each other out (if, for example, variety makes up for smaller volumes). The weight of each of these characteristics on the value of the data will also be determined by the end market in which the big data will be used, thereby creating different-sized economies for different markets.

The operative empirical question is, of course, at what point diseconomies set in, and under what conditions these barriers are sufficient to create durable and substantial market power. If the scale, scope, and speed of data collection at which diseconomies set in is sufficiently large, the cost of data collection can make entry prohibitive. One example in which economies of scale encompass the whole panel of relevant data involves patented goods. To avoid infringement of a valid patent, full information about all relevant patents is essential - partial data will not do.[51]In other cases, however, a much smaller dataset might be sufficient. Alternatively, the freshness of the data might give it most of its value,

economies of scale. *See, e.g.,* Michele Banko and Eric Brill, *Scaling to Very Very Large Corpora for Natural Language Disambiguation* (2001) http://research.microsoft.com/pubs/66840/acl2001.pdf; *See also,* Enric Junque de Fortuny, David Martens, and Foster Provost, *Predictive Modeling with Big Data: Is Bigger Really Better?* 1 BIG DATA 4 (2013). We are grateful to Hal Varian for pointing to these examples.

[48] *See* ALFRED D. CHANDLER, JR., SCALE AND SCOPE: THE DYNAMICS OF INDUSTRIAL CAPITALISM 281 (1990).

[49] Stucke and Grunes, *supra* note 19; Stucke and Grunes, BIG DATA, *supra* note 22, p. 8..

[50] *Id.*, pp. 285-87.

[51]*See, e.g.,* Lerner, *supra* note 17, at 36-7. Of course, the ease of determining whether an infringement occurs is also affected by the organization of the relevant data regarding patents.

thereby limiting the significance of other characteristics such as volume or access to past data, and potentially lowering barriers to entry.[52]

Observe that to achieve such economies the data need not relate to the specific final market. Rather, there might be externalities with regard to its analysis and application. For example, a health research supplier in one country (such as Kantar Health Inc.) might enjoy economies of scale, scope, and speed in a country in which it has never before performed research, so long as there are sufficient similarities in the data in both countries. Also observe that sometimes a difference in scale, scope, or speed can translate into a difference in kind- creating a different relevant market for the data.

Determination of the size at which diseconomies arise can be difficult, since it requires knowledge about the speed of the data being extracted along with its scale and scope. As an illustration, consider the ongoing debate about the presence of entry barriers with respect to search. Microsoft has argued that it faces substantial barriers to entry because it obtains an order of magnitude fewer search queries than does Google.[53] From Microsoft's perspective, its analysis of its own queries puts it at a disadvantage. Google counters by pointing out that efficient scale can be readily achieved through the analysis of queries on Bing, suggesting that if Microsoft is disadvantaged it is due to Google's more successful algorithm or other comparative advantages and not to scale.[54] This implies that different data analytical tools can create divergent economies of scale. The same may result from different quality of raw data collected: "dirty" or corrupt data (meaning data from which not much information can be easily gleaned) versus clean data. All these considerations must be taken into account when determining the magnitude of the various economies.

Of course, barriers created by economies of scale, scope, and speed are not necessarily insurmountable. Often-used examples are Google's displacement of Yahoo! as the primary U.S. search engine and Facebook's entrance into the social media space, largely replacing earlier entrants such as MySpace. Yet these examples are limited in their implications, since the scale of data collection is primarily affected by the quality of other services, such as search or social media. Indeed, Google's launch of free

[52] *See, e.g.,* Nils-Peter Schepp and Achim Wambach, *On Big Data and Its Relevance for Market Power Assessment,* JOURNAL OF EUROPEAN COMPETITION LAW & PRACTICE (2015).

[53]See the "Search" discussion at fairsearch.org.

[54] This leaves open the question of whether Bing remains at a disadvantage with respect to queries at the "tail" of the distribution.

Gmail created a comparative advantage.[55] Moreover, in both examples the entrants were able to substantially enlarge the scale of collected data relative to previous entrants. Therefore they only indirectly and partially signify the scale of data-collection economies, or the weight to be given to past data in such markets. Better examples involve purely data-collecting firms that compete intensively, despite scale differences among the competitors.

Finally, a data collector intermediary can sometimes assist firms in overcoming barriers such as sunk costs in data collection, by aggregating the demand for data over several consumers.

c. Network Effects

Data-driven network effects can create a demand-side technologically-based barrier to entry.[56]If the benefits that individuals receive are positively related to the number of other individuals that utilize or consume a product, the resulting barrier will have an effect that is similar in its impact to a more traditional supply side barrier. Substantial (sunk) expenditures will be required to counter or even overcome existing network effects. This may happen when the quality of the product depends on the quality of the data, which, in turn, is affected by the number of data entries, their variety and freshness. This is because such data accelerates automated learning. Entry of new firms which do not have such data might be quite difficult.[57] Indeed, since correlations are often data-driven rather than theory-driven, the more wide-scoped and verified the data, the better the information derived from it. As Stucke and Grunes put it, "the more people actively or passively contribute data, the more the company can improve the quality of its product, the more attractive the product is to other users, the more data the company has to further improve irs product, which becomes more attractive to prospective users."[58]

[55] For a similar conclusion see the French and German Report, *supra* note 14, p. 30.

[56] Network effects are present when the value of adopting a service to a new user increases with the number of users that have previously adopted the service.

[57] This was recognized, for example, by the European Commission in TomTom/TeleAtlas (Case Comp. /M.4854), Commission Decision C(2008) 1859 [2008]OJ C 237/12.p. 25; US Department of Justice, Statement of the Department of Justice Antitrust Division on its Decision to Close Its Investigation of the Internet Search and Paid Search Advertising Agreement Between Microsoft Corporation and Yahoo! Inc., 18 February 2010. See also Stucke and Grunes, BIG DATA *supra* note 22, ch. 12.

[58] *Ibid*, p. 170.

Social networks provide a classic example. Facebook users benefit from having a large group of "friends" who belong to the same network,[59] and the same is true with regard to network-based services, such as TripAdvisor or Yelp. Unlike the supply side, with demand-driven economies the share (rather than size) of the market can be crucial. To illustrate, the more data about the quality of hotels based on reviews from past users can be found on Tripadvisor, the more valuable the data-based information to each user. This, in turn, can erect barriers into the market for the collection of data (e.g., creation of a new competitor to Tripadvisor). Furthermore, a positive feedback loop can be created: as more users use the site, more information is gathered, and better targeted information (and advertisements) can be sent to users. The revenues from the paid side (advertisements) can be used to create a better service (creating a qualitative comparative advantage) which, in turn, attracts more users or incentivizes them to spend more time on the service's site, which in turn generates more data that improves the service, and vice versa.[60]

The central empirical questions are: What is the minimum scale that makes entry viable in order to overcome such network effects? Is the minimum efficient scale achievable? As with supply side barriers, the answers to these questions depend not only on the cost of extracting or otherwise obtaining data, but also on the quality of the data analysis. It also depends on the unique characteristics of each market, and therefore may change from one market to another. Furthermore, a cost-based analysis of scale is often not pertinent. As an alternative, we can make inferences through revealed preference. To illustrate, we know with respect to social networks that Facebook faces competition from Google, Linked In, and other broad-based networks, as well as more focused competition from niche networks such as Wiser and Behance. Similarly, with respect to messaging and email, Google faces competition from Apple, Facebook, and Microsoft, among others. Yet here again the quality of the service which creates the platform for the data collection is an essential part of the analysis.

[59] Such network effects were recognized by the European Commission in Case Comp/M.7217 Facebook/Whatsapp, Commission Decision C(2014) 7329 final, 3 October 2014, para. 12. For an analysis of the decision see Stucke and Grunes, BIG DATA, *supra* note 22, at 165-69. Yoo argues that Facebook is not a good example of network effects. This is because the average Facebook user actively interacts with no more than a handful of people. Christopher S. Yoo, *When Antitrust Met Facebook,* 19 GEORGE MASON L. REV. 1147 (2012). Yet this argument disregards the fact that one's circle of Facebook friends often includes people who are part of other circles as well. Network effects in data-driven industries were recognized in *United States v. Bazaarvoice*, Case No 3:13-cv-00133-WHO, 2014 WL 203966 (US Dist Ct (ND Cal) 8 January 2014, slip op, p. 24.

[60] *See also* Stucke and Grunes, BIG DATA, *supra* note 22*,* p. 189-99.

d. Two-Level Entry

In some instances data collection is performed as a stand-alone action. The Thomson/Reuters proposed merger provides an interesting example.[61] As described by Robert Mahnke,[62] both companies offered, among other things, bundles of financial data to traders and other financial professionals. The DOJ, as well as the Competition Directorate of the EU, found that there was a barrier to entry with respect to fundamentals data for publicly traded companies worldwide. Specifically, the DOJ alleged that new entrants into the international side of the fundamentals data market would have difficulties arranging for the collection of data on tens of thousands of companies on a global basis, constructing a reliable historical database, developing local expertise in each country's accounting norms, and developing data normalization and standardization processes. “Therefore, entry or expansion by any other firm will not be timely, likely, or sufficient to defeat an anticompetitive price increase.”[63]

More commonly, data are collected as a (valuable) side-effect of other productive activities. To illustrate, data regarding geological conditions are often a by-product of deep drillings in search of underground valuable resources. Likewise, scanner data, which are an important input in the market for electronic market-based tracking services, are a by-product of supermarket sales of goods. Or data gathered by doctors on how their patients react to a certain medicine for a long-run study, are in each individual doctor’s point of view, an important but still marginal activity. Each of these datasets is difficult to replicate by another firm which is interested in the data rather than in finding oil, selling grocery goods or administering medicine. In each instance, a two-level entry barrier for the collection of data arises, should the data be unique and not easily replicable, or if data portability might be limited due to technological, legal or behavioral barriers.

e. Two-Sided Markets

Two-sided markets are characterized by a platform which acts as an intermediary between two groups of users, whose demands are interdependent, therefore creating cross-platform externalities that the

[61] Cases in which data were bought and sold include, for example, Thomson/Reuters (financial information) (U.S. v. Thomson Corp. and Reuters Group PLC, Case No. 1:08-cv-00262 (D.D.C. 2008) and Thomson Corp./Reuters Grp, EU Comm’n Decision (Feb. 19, 2008)).

[62] Robert P. Mahnke, *Big Data as Barrier to Entry*, CPI ANTITRUST CHRONICLE May 2015 (2) (While the paper focuses on traditional data market, its analysis has broader implications.

[63] U.S. v. Thomson, *supra* note 61. *See also* European Comm’n Decision, NewsCorp/Telepiù, Case COMP/M.2876 [2004] OJ L 110/73, ¶¶ 361-68.

intermediary seeks to exploit.[64]This type of barrier arises especially with relation to the collection of data regarding users' actions that reveal preferences. Since the wealth of on-line options has created a poverty of attention,[65] firms often compete over users' attention (competition over eye-balls), in exchange for the ability and right to harvest and use data regarding users' on-line actions.[66]For example, it is commonplace to offer on-line services free of direct charge (e.g., automatic translations, currency exchange rates, etc.) in exchange for the (indirect) disclosure of data, which is then often used as an input for monetization services. This, in turn, implies that entry into data-collection activities in such markets is affected by the ability to provide competitive services or goods that lure consumers to these on-line services.[67]

Search engines provide a prototypical example. Search data are highly valuable; they allow Google, Microsoft, Yahoo!, and others to serve a matching function – to act as a go-between putting two distinct user groups – the advertisers and the sites that place the ads. In order to obtain such data, search companies often provide valuable search services free of direct charge. The higher consumers' switching costs (e.g., because of stored preferences), and the more important search engines or on-line services become gateways to content,[68]the higher the barrier for competitors to collect data.

Users therefore play an indirect role in the erection or increase of entry barriers in data-collection markets. Their choice of which products and services to use is affected by a combination of parameters, including users' (mis)information regarding the quality and price of competing goods or services from which the data are collected, switching costs, and (mis)information or (in)difference about the indirect price they pay in terms of lost privacy, in objectification, in the right to be forgotten, or in the erection of entry barriers. As we elaborated elsewhere,[69] the welfare effects of such barriers may be substantial.

[64] *See, e.g.*, Marc Rysman, *The Economics of Two-sided Markets*, 23(3) JOURNAL OF ECONOMIC PERSPECTIVES125 (2009); David S. Evans and Richard Schmalensee, *Industrial Organization of Markets with Two-Sided Platforms*, 3 COMPETITION POLICY INTERNATIONAL 154 (2007); Mark Armstrong, *Competition in Two-Sided Markets*, 37(3) RAND JOURNAL OF ECONOMICS 668 (2006).

[65] Simon, *supra* note 7.

[66] Evans, *supra* note 6.

[67]See Evans and Schmalensee, *supra* note 64; Gal and Rubinfeld, *supra* note 8.

[68] Nico Van Eijk, *Search Engines, the New Bottleneck for Content Access*, in TELECOMMUNICATION MARKETS: DRIVERS AND IMPEDIMENTS (Preissel, B., Justus Haucap and Peter Curwen eds., 2009)(search engines are the most important gateway used to find content).

[69] Gal and Rubinfeld, *supra* note 8.

Two-sided markets do not, however, necessarily imply that entry barriers are high. To illustrate, evidence of multi-homing is suggestive (but not dispositive evidence) of the presence of low to moderate switching costs.[70] The ability of a meaningful number of users to alter the mix of usage of two or more data-collection sources might imply that entry barriers are not prohibitive.

Observe that two-sided markets create an interesting dynamic with regard to the scale, scope, and speed economies of big data. This is due to the fact that the big data's quality (as affected by the size of its economies) might have a different effect on each of the two sides of the market. This, in turn, implies that in the analysis of entry barriers into two-sided the markets, the quality of the big data must be combined with other parameters. Search engines provide a good illustration. The quality of the data collected on users directly affects its value for advertisers and website operators. Yet the value of more personalized information based on big data analysis might vary among the search engine's users (possibly even creating a negative externality for some) and among types of information (a targeted ad might be worth less to a consumer than a targeted newsfeed). Furthermore, oftentimes the value to the user will be affected by parameters that do not involve big data (e.g., the aesthetics or ease of use of the website, etc.). Moreover, the data collected on a certain user might not necessarily be used (solely or at all) to create better services for himself, but might benefit other users, thereby creating one-step-removed effects (this might happen, for example, when the data analysis is not immediate and will only benefit future users).

Diagram2 illustrates some of the effects of the quality of data (as determined by its economies) in two-sided markets. Data quality creates direct effects (narrow arrow) as well as indirect effects (bold arrows). For example, the quality of the data directly affects the ability of the advertiser to create a more targeted ad to each specific user. This, in turn, might have indirect effects on the user (unceasing his incentive to buy the product offered, or creating unease him because of a tracking feeling) as well as on future users. These effects on users might, in turn, indirectly affect the advertiser (if, for example, the user might choose to use a less intrusive online service) All these factors combined create complicated dynamics with regard to the optimal volume, variety, velocity and veracity of the big data to be collected, as well as to its competitive effects on the market. This dynamic may explain the fierce competition that exists in some on-line

[70] See, *e.g*., Aaron S. Edlin and Robert G. Harris, *The Role of Switching Costs in Antitrust Analysis: A Comparison of Microsoft and Google*, 15 YALE J. L. TECH.169 (2013)(Suggesting that in online markets, costs to users of switching platforms or multi-homing are very low).

markets and the entry of new firms despite the fact that they enjoy less specialized or lower quality or quantity of data.

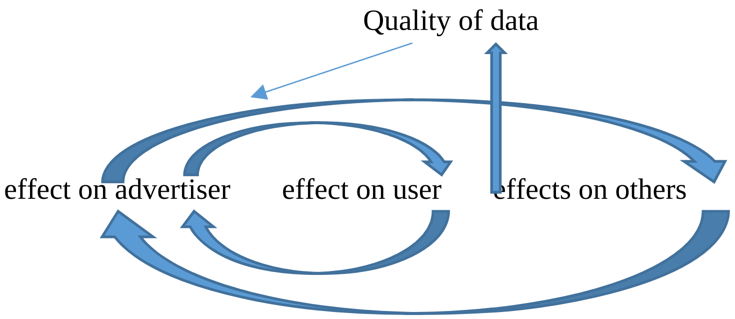

Diagram 2: Direct and indirect effects of the quality of data

f. *Information barriers*

Barriers to data collection might also arise from limited information on who owns the relevant data, or on the costs of locating and contracting with such data holders. For example, should data describing how patients react to a certain medical condition be located in the files of many doctors, all over the world, difficulties may arise in locating and gathering such data. Observe, however, that in some fields the government reduces information costs. For example, patent offices normally register and publish information regarding all valid patents in their jurisdiction.

2. Legal Barriers

Legal barriers play an ever-increasing role in data collection barriers. While legal barriers are often justified by broader goals (e.g., privacy, limiting objectification and discrimination), they also carry costs in the form of limiting access to data. The height of these barriers will be affected, inter alia, by the scope of the legislation as well as whether it embodies property or liability rules.[71]Legal barriers create direct as well as indirect barriers to the collection of data (either self-collection or transfer from another data collector). The latter category includes self-imposed limitations on data portability. These are based on concerns that once the database is revealed to other entities, the data-gatherer will be more exposed to claims that it infringed its legal obligations in its data collection activities.

Below we explore legal barriers that directly apply to data collection. We note, however, that additional ones may apply indirectly. For example, copyright limitations in certain jurisdictions might limit the provision of certain content, which then serves as a basis for data-collecting activities (e.g., Netflix, Spotify).

a. Data protection and privacy laws

An increasing number of jurisdictions have imposed limitations on data collection activities, most commonly with regard to personal data, [72] in order to protect privacy as well as other social goals, such as non-

[71]Guido Calabresi and Douglas Melamed, *Property Rules, Liability Rules and Inalienability: One View of the Cathedral*, 85 HARVARD L. REV. 1089 (1972).

[72] In the EU privacy is treated as a fundamental right. For protection of private data in the EU *see, e.g.*, EU Data Protection Directive 1995 http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:31995L0046; Draft of General Data Protection Regulation, http://ec.europa.eu/justice/data-protection/index_en.htm; Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), Sections 22, 27-29 (limiting the storage of personal data) http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML.Protection of databases can be done through different legal means. *See, e.g.*, Directive 96/9/EC of the European Parliament on The Legal Protection of Databases, http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009 (protection applies if the database is the result of substantial investment in obtaining, verifying or presenting its contents). Copyrights can also protect the whole database if it is an original literary work. The test of originality requires that effort be spent on the selection and arrangement of the data and sufficient judgment and skill exercised in the process to making of the database. The U.S. does not have a coherent single legislation regarding the protection of data and databases; rather, there is complex "patchwork system" of federal and state laws. It is interesting to note, that economic studies have shown that no conclusive conclusions can be drawn about whether privacy protection is beneficial to society. *See, e,g,* Wolfgang Kerber, *Digital Markets, Data and Privacy*, GRUR INT. 639, 641 (2016); Alessandro Acquisti, Curtis Taylor and Liad Wagman, *Economics of Privacy*, J. ECON. LIT. 2016 (forthcoming).

discrimination of school children.[73]While these laws create barriers for data collection, they can often be overcome by ensuring the anonymity of the data.

When the law blocks a method of collection, it creates a need to find an alternative technological solution, thereby erecting entry barriers at least for some firms. One interesting example involves the use of “cookies” as a means of collecting information. Despite their cute name, tracking cookies are technological devices that allow website owners to expand their data collection to activities of the users on other websites, by inserting links to databases. While in the U.S. the use of cookies is not prohibited, the EU has recently adopted limitations on their use. Under EU regulation, the user must give permission to the use of cookies in each and every site he enters (opt-in mechanism), thereby creating a legal barrier for data collection.[74] The goal of the regulation is to protect the autonomy and privacy of users-enabling them to set some limits on the use of the data collected about them.

This is an informative example for several reasons. First, this is, at least partially, a paternalistic law, based in part on the misinformation, misunderstanding or apathy of many users with regard to the implications of data collection on their individual and collective well-being.[75]Second, its justification is also based on the high switching costs consumers sometimes face in using products and services that do not infringe on their privacy (e.g., DuckDuckGo[76]). Third, and most importantly, it creates indirect competitive implications. The limited use of cookies to gather data gives Google a comparative advantage over its potential and existing rivals because it uses another technological route for data collection (mostly through its search engine).Since Google holds a strong position with respect to search in the EU, it can use this platform to collect vast amounts of data without the use of cookies.[77]This is not to say that privacy-based limitations on data collection do not serve the public interest. Indeed, if there are anticompetitive effects, they may well be offset by wider social welfare considerations. Either way, a first step in an appropriate welfare

[73] Of course, much depends on how private data is defined. Certain types of data are further protected. For example, certain data regarding schools is protected by the U.S. through the Safe Kids Act 2015, S. 1788.

[74] Regulation (EC) No 2006/2004 on consumer protection cooperation OJ EU L 337 of 18 December 2009, p. 11. The directive has been adopted so far by all European Union Member States.

[75]For example, a survey conducted by the Pew Research Center in 2014 found that half of online Americans do not know what a privacy policy is. Pew Research Center, What Internet Users Know About Technology And the Web (2014), http://www.pewinternet.org/files/2014/11/PI_Web-IQ_112514_PDF.pdf.

[76]Privacy Policy, DUCKDUCKGO.COM, https://duckduckgo.com/privacy (last visited 17 April, 2016)(no personal data collected).

[77]Google does not require individuals to disclose personal data.

analysis is to recognize the entry barriers created by the regulation, which is the focus of our analysis.

b. Data ownership[78]

Another set of legal barriers arises from data ownership issues, which affect the ease of access to the data despite its non-rivalrous nature. To illustrate, the identity of the one who owns the data regarding a person's medical history, which might be an essential input for future medical treatments, may determine the height of a barrier to entry into the medical services market.

While no specific law grants ownership to big data as such, some forms of data collections are protected. Legal systems generally differentiate between "raw data" and databases. Raw data refers to basic, unprocessed data such as Internet traffic. Generally, raw data, including private data, are not seen as owned by anyone. Furthermore, as the recent EU *Schrems* case indicates, collecting data using one's own algorithms does not automatically make the collected data the property of the collector.[79] The protection of databases is elaborated below, in the section about usage.

3. Behavioral Barriers

Consumers' behavioral biases often serve to reduce entry barriers. Their ignorance or indifference towards the collection and use of their personal data lowers potential entry barriers. Below we explore several examples of another type of behavioral barriers to data collection: those erected by data collectors.[80]

a. Exclusivity

Contractual exclusive access to a unique source of data may create entry barriers in the form of input or outlet foreclosure.[81] The Canadian Nielsen case provides an illustrative example. Nielsen competed in the market for electronic market-based tracking services, which was based on scanner-data. It entered into exclusive contracts with all the major supermarkets in

[78] See, e.g., Josef Drexl, *Designing Competitive Markets for Industrial Data – Between Propertisation and Access* (2016) (on file with author)(analyzing, inter alia, the current focus of the EU on the creation of a single digital market)

[79] European Court of Justice, Case C-362/14, Maximillian Schrems v. Data Protection Commissioner [not yet published], http://curia.europa.eu/juris/liste.jsf?num=C-362/14.

[80] Additional examples exist, such as increasing user's switching costs, or leveraging data advantages from a monopoly position in a regulated market to other markets. *See, e.g.*, BIG DATA, *supra* note 22, chapter 18.

[81] The Department of Justice recognized this concern in the Google/ITA software case: United States v. Google Inc. Case No 1:11-cv-00688 (US Dist Ct (D DC), 8April 2011), Competitive Impact Statement.

Canada regarding their scanner data.[82]This effectively excluded its rivals from the market. It is noteworthy, however, that due to the data's non-rivalrous nature and the fact that consumer preferences can sometimes be observed from alternative sources of data, exclusivity over one source of data often may not create a significant barrier. *Access Prices and Conditions*

Another behavioral access barrier involves access prices and conditions set by the data owner for granting access to his data. These terms need not necessarily relate to the effort invested in their collection (or organization and analysis), to their market value, or their value to the one requesting access. Once again, they will be prohibitive only if the dataset is unique and cannot be easily or cheaply replicated.

b. What to Collect

Even if a firm has unique data collection abilities, it might not be in its interest to collect certain kinds of data which might give its rivals a comparative advantage. An interesting example- albeit one that is based on political motivations- involves census data collected by the South African government. Following a heated public discussion on a question pertaining to the religious beliefs of the population, the South African government decided to take the question off the census.[83]

c. Disabling Data Collecting Software

It is commonplace for firms, especially browser owners, to disable one another's data gathering mechanisms. For example, each update of Microsoft's operating system erases the search engine currently in use, and sets Microsoft's own browser (Bing) as the default, although the browser is not part of the operating system update.

C. BARRIERS TO THE STORAGE OF BIG DATA

1. Technological Barriers?

Storage is one of the main areas in which technological advancements have significantly reduced entry barriers to big data markets.[84]Storage has the potential to create three types of barriers. The first involves storage space. In the past, if a firm did not have sufficient space on its hardware to store its data, it faced a high technological barrier. Advancements in both

[82] Investigation & Research v. D&B Co. of Canada, Reasons for Order, Competition Tribunal, 1995.

[83] ALEXANDER JOHNSTON, SOUTH AFRICA: INVENTING THE NATION 234 (2014).

[84] EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES1 (2014).

hardware and software have substantially reduced his problem. Computers have today much better storage capabilities than before. But more importantly, the hardware problem was largely overcome by a software solution: the creation of cloud computing and the strength of the Internet have served to create a system where firms can rent computer storage space. Moreover, firms can upload and download data from anywhere to everywhere, as long as they have a prior agreement to use other firms' storage capacity. Yet in some instances storage might still be an issue.[85]The second barrier, that was also substantially reduced, involves energy-driven storage costs. New advancements in data storage have served to reduce these costs significantly. A final barrier involves software for efficient storage. These costs have also been steadily declining. Indeed, currently there is free open-source software for storage and access (Cassandra, HBase).

2. Lock-in and Switching Costs

Probably the largest barrier to competition in data storage is lock-in and switching costs. As Shapiro and Varian have observed, "[s]witching costs and lock-in are ubiquitous in information systems."[86]Indeed, in some of our interviews big data companies reported high switching costs due to storage. This is because once the data are stored in a particular order, it may be difficult to transfer elsewhere, especially if the storage order is not known to the user, or is protected by law. Switching is also likely to be difficult if the software necessary to manage the data varies widely between datasets. Suppose, for example, that a firm is using HP data management services for its vital business data. A switch to an Oracle-based system would likely entail substantial costs. This implies that a choice of database management software might create high switching costs that sometimes lead to lock-in effects.

3. Legal Barriers

Recently, a legal barrier was erected in the EU with regard to the storage of data elsewhere. An Austrian citizen, Max Schrems, challenged the transfer of his personal data by Facebook Ireland to be stored in Facebook USA.[87] The European Court of Justice invalidated the 'Safe Harbor' agreement between the EU and the U.S. in which the transfer of

[85] Server capacity might be an issue in indexing the web. Google is using more than 1 million computers to index the web. *See, e.g.*, http://www.extremetech.com/extreme/161772-microsoft-now-has-one-million-servers-less-than-google-but-more-than-amazon-says-ballmer

[86] Shapiro and Varian, *supra* note 27.

[87]Schrems, *supra* note 79.

personal data was allowed, based on privacy considerations. This judgment is expected to have implications for international businesses that market their products to the EU, with regard to the location of their data, as well as their ability to integrate data collected in the EU with other sources of data.[88]

D. BARRIERS TO SYNTHESIS AND ANALYSIS OF BIG DATA

Many of the barriers that characterize the collection of data, such as scale and scope economies, may be relevant to data synthesis and analysis. Below we analyze two additional barriers, which are unique to this stage in the data value chain, both of which are technological barriers.

1. Data Compatibility and Interoperability

Each data owner organizes his data in a way which fits his own needs and preferences. Such organization might create a barrier for synthesis with other data or for its use by others. Several barriers are relevant. First, one should know what each rubric in the dataset stands for and how exactly it was measured, in order to determine its relevance and trustfulness. To illustrate, data relating to public transportation schedules may lead to different results for determining efficient traffic routes, depending on whether the data relate to scheduled or to actual times of arrival. Indeed, there is often no transparency in data organization to an external observer; rather, the dataset is a "blackbox," unless the data organizer shares his data-about-data, a fact which limits data portability and interoperability. Second, barriers may arise from the way the data are organized, even if all parameters are known. This is especially problematic where the database includes numerous parameters or when it is constantly updated.

2. Analytical Tools

The availability and quality of algorithms used for synergy and analysis of big data may create a technological barrier. While some analytical tools are available free for everyone to use, much depends on the level of analysis required. Since most of the raw data is unstructured (e.g., words, images and video), the tools for gleaning knowledge play an important role. These include, for example, rapidly advancing techniques of artificial intelligence like natural-language processing, pattern recognition, and machine learning. The differences in data science tools between firms (in the efficiency of the analysis both energy-wise and decision-quality-wise) might create comparative advantages for some firms. Indeed, some big data analyzers, such as Google, invest large sums of money in developing or in

[88] Following the decision, the EU and the US reached an agreement on transatlantic data flows. The new arrangement will provide stronger obligations on companies in the U.S. to protect the personal data of Europeans and stronger monitoring and enforcement by U.S. authorities. *See* http://europa.eu/rapid/press-release_IP-16-216_en.htm.

buying advanced algorithms.[89] Such advantages might be especially important where data synergies are important. At the same time, steadily declining costs of computing, including memory, processing, and bandwidth, imply that the use of data-intensive analytical tools are quickly becoming economical.[90]

Observe that a correlation may exist between the quality of the data and the quality of the algorithm, due to the algorithm's feedback loop and potential future changes in the algorithm that evolve from data scientists' observations regarding past corrections in the data.

E. BARRIERS TO THE USAGE OF BIG DATA

Even if one possesses or can access big data through data intermediaries, this does not automatically imply the data can be effectively utilized.[91]

1. Technological Barriers

Technological barriers may involve the inability to locate and/or reach relevant consumers. Assume, for example, that a firm possesses first-rate information based on its unique dataset. Yet it cannot reach the relevant consumers. This might be the case if the platform that consumers use is controlled by one of its rivals. Alternatively, the data-owner cannot easily locate consumers who might benefit from its dataset. Interestingly, this latter obstacle can sometimes be overcome by another dataset, which focuses on consumers' location and preferences.

2. Behavioral barriers

Self-imposed contractual limitations on the ability of the data collector to use the data for certain uses or to transfer it to others may create barriers to the firm's own use or to data portability. These limitations usually apply to private data. They are used by firms in order to increase usage of their data-generating products by creating a reputation as private-data protectors,[92] or

[89] It is interesting to note that such mergers and acquisitions are often not scrutinized under merger review, especially if the new technology was not used in the market when it is bought and therefore did not already create market power. In the European context see e.g., Inge Graef, *Market Definition and Market Power in Data: The Case of Online Platforms* 38(4) WORLD COMPETITION: LAW AND ECONOMICS REVIEW (2015).

[90] Andrew McAfee and Erik Brynjolfsson, 'Big Data: The Management Revolution, HARVARD BUSINESS JOURNAL (2012), https://hbr.org/2012/10/big-data-the-management-revolution/ar.

[91] See also French and German Report, *supra*, note 14, at 41.

[92] Infringement of such obligations can amount to both contractual violations as well as violation of consumer protection laws. See, e.g., FTC, Press Release, FTC Approves Final Order Settling Charges Against Snapchat (2014), https://www.ftc.gov/news-events/press-releases/2014/12/ftc-approves-final-order-settling-charges-against-snapchat (based, inter alia, on allegations that Snapchat deceived consumers about the "amount of personal data it collected.").

catering to a segment of the market which values privacy. They might also be used in order to limit pressures to share data or to prevent multi-homing. An interesting example involves the U.S. government's request to access Apple's data in order to seek information on potential terrorist acts. Apple refused, based on self-imposed privacy concerns.[93]

Another type of contractual restraints involves limitations that data collectors impose on user's data portability. Such restrictions limit the user's ability to export personal data from one application to another without encountering too much difficulty. Such limitations increase switching costs and may generate lock-in effects. Known examples are the ability of competing advertising platforms to make use of exported Google ad campaign data,[94] and Facebook blocking a Google extension that would have allowed data to be exported from Facebook to Google+, a rival social network.[95] It is noteworthy that the European General Data Protection Regulation includes the right to data portability.[96]

The erection of such barriers is dependent, of course, on the incentives of big data owners to create them. This, in turn, is case-specific. For example, where a firm's business model with regard to data is based solely on data collection, storage, or analysis (for example, where the data is a side-effect of another productive activity and has limited use in that activity), the incentive to share it is high. The costs of entering another part of the value chain (e.g., analyzing the data collected) might also affect the incentive to share data. However, where a firm's comparative advantage relies on the use of a unique data set, and it is best positioned to make efficient use of this data, its incentives to limit the transferability of the data will be higher. Note, however, that the analysis of such incentives is no different than the traditional analysis with regard to the sharing of other inputs.

3. Legal Barriers

Often the main barriers to usage are legal limitations designed to protect users' privacy. The European right to be forgotten serves as an illustrative example.[97]Interestingly, the scope of such legal barriers is often informed

[93] USA v. In the Matter of the Search of an Apple iPhone Seized During the Execution of a Search Warrant on a Black Lexus IS300, California License Plate 35KGD203 (California Central District Court, 2016). The FBI eventually withdrew its request (March 28, 2016).

[94] FTC Press Release, Google Agrees to Change Its Business Practices to Resolve FTC Competition Concerns in the Markets for Devices Like Smart Phones, Games and Tablets, and in Online Search (2013), https://www.ftc.gov/news-events/press-releases/2013/01/google-agrees-change-its-business-practices-resolve-ftc.

[95] *See, e.g.*, Emil Protalinski, *Facebook blocks Google Chrome extension for Exporting Friends* (2011), http://www.zdnet.com/article/facebook-blocks-google-chrome-extension-for-exporting-friends/

[96] European General Data Protection Regulation, Section 18.

[97] Regulation (EU) No XXX/2016 of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing Of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), Article 17,

by technological capabilities: the ability to make the data anonymous may determine the scope of the data that can be legally used.

Laws that protect additional social goals may also erect barriers, either directly or indirectly. Equal opportunity laws, which prohibit discrimination based on certain characteristics, may erect an indirect barrier. For example, a lender cannot offer less favorable terms to single compared to married persons even if big data analytics show that the former are less likely to repay loans than the latter.[98] Similarly, data protection laws prohibit the sale of big data analytics products to customers if they have reason to know that the data will be used for fraudulent purposes.[99]

Intellectual property protection creates a direct barrier. As big data becomes more important, we can expect to see more issues arising with regard to the scope of intellectual property protection. One example involves the protection of databases,[100] which is currently debated in several EU countries. Questions being considered include: Should we protect the database even if not the raw data, and when a database should be protected. Oftentimes the answer depends on the level of expertise and work used to create the database.

An interesting example involves the Myriad/Ambry dispute. In July 2013, Myriad Genetics sued Ambry Genetics for patent infringement relating to genetic diagnostic testing.[101] Ambry had previously announced that it would provide genetic diagnostic testing for certain genes. In its complaint, Myriad pointed to its $100 million investment in developing its extensive database of genetic invariant information, which could ensure a high degree of accuracy in its testing.[102] In its Preliminary Injunction submission, Myriad claimed that Ambry was "free-riding" off its investment in creating the dataset genetic testing. Ambry countersued, claiming that Myriad was using its claim to monopolize the specific testing market. The district court in Utah denied Myriad's motion, relying on Myriad's inability to overcome a "substantial question" as to whether Myriad's patent claims were valid. An appeal was rejected by the Federal

www.janalbrecht.eu/fileadmin/material/Dokumente/GDPR_consolidated_LIBE-vote-2015-12-17.pdf.

[98] See FTC, Inclusion, *supra* note 32, at iii. In a separate statement, FTC Commissioner Maureen K. Ohlhausen argued that "to the extent that companies today misunderstand members of low-income, disadvantaged, or vulnerable populations, big data analytics combined with a competitive market may well resolve these misunderstandings rather than perpetuate them." *Ibid*, at. A-2.

[99] *Id*, at iv.

[100] EC Council Directive on the Legal Protection of Databases (1996).

[101] University of Utah Research Foundation, et al, v. Ambry Genetics Corporation, United States District for the District of Utah, Case No. 2:13-cv-00640-RJS.

[102] Foley and Lardner LLP, "Myriad's Trade Secret Trump Card: The Myriad Database of Genetic Variants" (2013), *available at* www.pharmaptentsblog.com/2013/07/18.

Circuit.[103] In many respects these issues are no different than the issues that arise in many patent infringement cases. The distinction here is that, if granted, the patent would grant a big data monopoly, a monopoly whose effects could well extend beyond the point at which a patent would expire. This raises an important policy question: When, if at all, should the public's access to the Myriad's "sequence data and interpretive algorithms" be regulated in the public interest.

PART III: SOME ECONOMIC IMPLICATIONS

The above findings have implications for the analysis of competitive conditions in big data markets which, in turn, affect social welfare as well as optimal regulatory policy. Accordingly, this part attempts to tie together the findings from the previous parts, to indicate how the characteristics of big data and the entry barriers at each level of the data value chain affect the competitive analysis.

A. ENTRY BARRIERS

We begin with seven observations regarding entry barriers. A first and basic observation is that barriers can arise in all parts of the data value chain. Such barriers may be technological, legal, or behavioral. Often there exists a combination of barriers in each part of the value chain as well as across parts, which has the potential to create a cumulative negative effect on competition. At the same time, high entry barriers in one part of the chain might at least be partly overcome by another part of the chain. For example, better algorithms can allow a firm to learn more from less data.

Second, while some barriers are unique to big data, others apply more broadly (e.g., barriers arising from two-sided markets or network economies).

Third, the existence and strength of such barriers may differ among markets for big data, depending on their specific characteristics. Therefore, in order to evaluate the importance of such barriers, one needs to understand the unique characteristics of the data which serve as an input for each specific market, as well as the ways in which the data-based information reaches the consumer. To give an example, the velocity of data necessary in order to determine optimal traffic routes at any given time is significantly different from the velocity of data required to determine past

[103] In Re BRCA1 and BRCA2-Based Hereditary Cancer Test Patent Litigation (Court of Appeals for the Federal Circuit, cases 2014-1361, 1366)(Dec. 17, 2014).

trends in the purchase of toys. Where economies of scope are significant, firms may benefit from a network of data sources, thereby possibly increasing entry barriers for those outside the network. Any analysis of big data markets without such groundwork- what Balto and Lane call "lazy talk"[104]- is therefore problematic.

Fourth, and relatedly, big data is non-rivalrous, and collecting it does not prevent others from collecting identical data by comparable or different means. This has led some to claim that low entry barriers exist in big data markets.[105] Our analysis of entry barriers challenges this assumption, at least in some big data markets. This is because data gathering is only part of the data value chain, because there may still exist high entry barriers in the collection of some types of data, and because of the cumulative effect of entry barriers in several parts of the data value chain, each of which, on its own, might not seem to create a high entry barrier. The focus of analysis should therefore not be limited to the data collection stage (unless it is the only relevant activity).

Fifth, in some markets relevant data-based information can be based on substitutable sources of data, thereby increasing the ability to compete in its collection. The example of data on the average speed of drivers on dark streets exemplifies this point. The same is true with regard to storage, synthesis and analysis tools.

Sixth, the strength of entry barriers at any part of the value chain might affect the strength of related barriers. For example, the lowering of one barrier might affect the incentives of firms to erect higher barriers in another part of the value chain, in order to protect their data-based advantages. Alternatively, if high and durable technological entry barriers exist in one part of the chain, we might more easily justify imposing legal barriers of lower or similar height, on the same part of the chain, that would protect privacy or other social goals.

Seventh, firms may use a variety of strategies in order to erect entry barriers, some observable and some less observable, some objectionable and some less objectionable.

B. EFFECTS ON COMPETITION AND WELFARE

The above observations, combined with the unique characteristics of big data, allow us to reach some conclusions with regard to the competitive conditions in big data markets. Entry barriers in such markets may create competitive effects which are similar to those created by traditional goods (e.g., foreclosure, market power), and firms enjoying data-based advantages will be motivated to engage in exclusionary conduct and erect

[105] *See. e.g.,* Tucker and Wellford, *supra* note 15, at 3; Lerner, *supra* note 17.

[104] David A. Balto and Matthew Cameron Lane, *Monopolizing Water in a Tsunami: Finding Sensible Antitrust Rules for Big Data* (2016) ssrn.com/abstract=2753249.

artificial barriers to entry in order to maintain or strengthen their advantage, just as in any other market.[106] Yet, as we show below, the nature, scale, and scope of such competitive effects are affected by the unique characteristics of big data markets, which create interesting twists on the regular analysis, which may, in turn, affect theories of harm. We also show that due to the unique characteristics of at least some big data markets, the mere existence of high entry barriers into these markets, by itself, does not automatically lead to the conclusion that social welfare will be harmed.

1. Data are multidimensional

As observed above, the quality and value of data are affected not only by its volume, but also by its velocity, variety, and veracity. As a result, once one characteristic of big data exhibits high entry barriers, another characteristic might grow in importance in order to overcome the competitive advantages created by the first. For example, where past data are not easily available (therefore reducing the volume or temporal variety of data available), veracity or variety might gain importance in order to create a higher level of predictive certainty based on a smaller data panel.[107]Firms might also invest more resources in creating better analytical tools rather than in gathering more data. Therefore, in analyzing the competitive effects of entry barriers in big data markets, it is essential that one explore alternative routes of reaching data-informed conclusions, based on the data's different characteristics. Failure to account for this fact may imply that potential implications for innovation (i.e., dynamic efficiency) will be disregarded.

The issue can be illustrated with respect to the U.S. DOJ/FTC 2010 Horizontal Merger Guidelines.[108] The Guidelines make clear that the U.S. agencies are concerned with mergers and acquisitions that are likely to increase barriers to entry created by the presence of scale economies, but there is no mention of similar barriers that could be created if there are substantial economies of scope or of speed.

Another implication of the multidimensional characteristic of big data, which may have opposite competitive effects than the previous one, is that data from different sources can create important synergies.[109] While entry

[106] See also Grunes and Stucke, *supra* note 19, at 11.

[107] See Ganit Ashkenazi, *Encouraging competition between Data Intermediaries: A Big Data Harms Accelerator?* (on file with authors).

[108]U.S. DOJ/FTC 2010 Horizontal Merger Guidelines.

barriers generally do not prevent own-use synergies, they might prevent synergies between different data owners, for example by limiting data portability. This, in turn, may affect welfare, depending on the importance of the synergies to the quality of the data-based decisions. To illustrate, consider the analysis of the reactions of patients to a certain medical treatment, which, only if gathered across a very large panel of data gathered from many doctors, can generate essential insights for better care. In such situations the ability to share data within and possibly across health-maintenance organizations can substantially increase welfare.

Synergies are often technologically relatively easy to create among datasets, given that data are non-rivalrous and can often be easily replicated at low costs. Yet for synergies to take place, two conditions must be met: (1) The relevant parties must be aware of possible synergies ("the informational obstacle"); (2) All relevant parties must find it worthwhile to invest in the realization of potential synergies (“the motivation obstacle”).[110]

Big data may involve both types of obstacles. In particular, the fact that each data owner organizes its data in accordance to its own preferences, might create obstacles even if access to the data is allowed. But more importantly, the motivational problem might be affected by the technological, legal, or behavioral barriers (e.g. contractual limitations on data portability) observed above. This creates an important policy concern – how can we design regulatory rules and institutions and evaluate mergers and acquisitions so that socially-desirable synergies will be created.

In fact obstacles to achieving these synergies produce an anti-commons problem, in which goods controlled by more than one right holder are underutilized.**[111]**With respect to big data, the problem arises because even if the data owner can give access to its database and will have an incentive to do so if such conduct benefits him economically, barriers to achieving potential synergies might remain.

Any synergies that can be created through the existence of big data must be balanced against market power concerns. Such considerations were

[109] Grunes and Stucke, *supra* note 19 at 11 note: “efficiencies claims were made and considered in Microsoft/ Yahoo!, *United States v. Bazaarvoice*, and TomTom/Tele Atlas. In each of these matters, the parties claimed that the merger would allow a company to produce better products faster because of data.”

[110] Based, in part, on Michal S. Gal, *Viral Open Source: Competition vs. Synergy*, 8(3) JOURNAL OF COMPETITION LAW AND ECONOMICS 469 (2012).

[111]The argument was first developed in the context of a patent regime- the claim was that dispersed ownership of patents harms social welfare because it limits synergies. Michael A. Heller and Rebecca Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE (1998); Michael Heller, THE GRIDLOCK ECONOMY – HOW TOO MUCH OWNERSHIP WRECKS MARKETS, STOPS INNOVATION, AND COSTS LIVES (2008).

taken into account, for example, in the FTC's evaluation of the proposed merger of Nielsen and Arbitron. Nielsen has long been active in the provision of various television (TV) audience measurement services to content providers and advertisers. Arbitron had been the leading provider of radio measurement services. Each separately had expended substantial resources in the development a panel of individuals who served as the source of TV and radio ratings, respectively. Each had "the most accurate and preferred sources of individual-level demographic data for [television and radio] audience measurement purposes."[112]In recent years both Nielsen and Arbitron had made efforts to expand their services in the "cross-platform" arena. Cross-platform audience measurement services report the overall unduplicated audience size (i.e., reach) and frequency of exposure for programming content and advertisements across multiple media platforms, with corresponding individual audience demographic data. The FTC had concerns about competition within the market for cross-platform ratings.[113] The FTC claimed that Nielsen and Arbitron were the best-positioned firms to develop cross-platform services in the future, because of their large, representative data panels, which created synergies (data economies of scale and scope). In settling the case, Nielsen agreed to divest and license assets and intellectual property needed to develop national syndicated cross-platform audience measurement services. Nielsen also agreed that ownership of all the data generated by the Arbitron Calibration Panel would belong to another rating firm, while Nielsen would retain ownership of the data from the Arbitron general panel.

A third and final implication of the fact that the data is vast and multidimensional, is that it strengthens the possibility that the dataset could be valuable to many different users, operating in unrelated and distinct markets. This is further strengthened by the fact that recent advancements in data science have led to the development of algorithms that can engage in deep learning, whereby the algorithm seeks correlations in the dataset without specific directions. This, in turn, enables the data analyzer to find correlations that otherwise would not be explored. To give but one example, deep learning with respect to on-line transactions found a correlation between people who engage in on-line fraud and the use of capital letters in writing on-line messages.[114]Such information is relevant to

[112]Analysis of Agreement Containing Consent Order to Aid Public Comment 2, Nielsen Holdings N.V. & Arbitron, Inc., FTC File No. 131-0058 (Sept. 20, 2013).

[113]FTC, Complaint, In the Matter of Nielsen Holdings and Arbitron PLC, Docket No. C-4439 (February 28, 2014).

[114]Many thanks to Sagie Rabinovitz for this example.

both enforcement agencies as well as to private companies operating in a wide range of on-line markets. This characteristic of the data affects its value, which in turn affects incentives to collect and analyze it.

2. The Non-Rivalrous Nature of Data

We now move to an exploration of the competitive effects of big data being non-rivalrous, since it can easily and cheaply be copied and shared, at least technologically. In this case, data collectors and analyzers have the potential to sell or license their datasets to multiple users. Yet legal and technological barriers in all parts of the data value chain may limit data portability. For example, legal limitations on data portability based on privacy concerns, or technological data compatibility limitations, might reduce the potential benefits that are available due to the non-rivalrous characteristic of the data. With or without these potential barriers, there are likely to be strong economic incentives to maintain control over large datasets and/or to create structural barriers, potentially rendering at least parts of the chain non-competitive.

If the source of the barriers is inherently structural, and sharing the data is socially beneficial, a regulatory solution may be appropriate, perhaps by requirements that the data be made widely available at a reasonably and non-discriminatory cost. A potentially instructive model might be FRAND (fair, reasonable, and non-discriminatory) licensing rate requirements that are central to the operation of standard setting organizations.[115]Of course, any regulation should be sensitive to the fact that relative advantages of big data are often nuanced and complex. It is also important to stress that the data being non-rivalrous does not alter the fact that its collection, organization, storage, or analysis generally transforms it into a private good.

Another notable implication of the non-rivalrous characteristic of big data is that the comparative advantages resulting from a unique dataset might be short-lived. Assume, for example, that a unique panel of data regarding the geological conditions in deep drillings is used by an insurance company to more accurately calculate the risks associated with insuring drilling operations. This, in turn, might lead the data-holder to set lower insurance prices that reflect reduced risk. Other insurers, acknowledging the fact that the price reducer has access to the relevant

[115]*See, e.g.*, Jorge L. Contreras, *A Brief History of FRAND: Analyzing Current Debates in Standard Setting and Antitrust Through a Historical Lens,* 80 ANTITRUST L.J. 39 (2015).

data, might follow suit and reduce their prices, without ever observing the dataset itself. The increased use of algorithms to track price changes of competitors may further limit the life of a data-based comparative advantage. This reduces the market value of the data, as well as the incentive to collect and analyze it in the first place. Note that no intellectual property rights are infringed, because the copying relates to the results of the analysis on actions, rather than to the copying of the database itself. Another example involves content scraping: practices that make use of data collected by competing firms, such as scraping journalistic information or consumer reviews by adding them to one's webpage. Here, however, intellectual property protection might kick in.

Indeed, due to the non-rivalrous aspects of the dataset, free riding is likely to be an issue. Free-riding can broaden the benefits flowing from big data, but it can also reduce reward and therefore the incentive to invest in creating the database in the first place. Assume, for example, that a firm enjoys monopoly power over the collection and analysis of a certain kind of big data, and that the data enable the firm to create better-targeted offers to consumers. If these consumers seek better offers, they may share the information regarding the first firm's offer with the firm's rivals, thus enabling the latter through free riding to base their offers on second-hand information. Alternatively, if the first offer is made on-line and is widely observable, successive suppliers can rely on the publicly available information to create a better one of their own. In today's world, when the calculation of most on-line offers is made by algorithms which can observe and analyze information regarding competitors' offers in split seconds, the comparative advantage of the first data-based offer might be reduced.

Free riding thus creates mixed effects on social welfare. The tension between the advantages of free-ride (greater competition, synergies) and the disadvantages (reduced incentive to innovate, market power) is inherent. This raises both regulatory and competition enforcement issues. Below we suggest some additional factors that affect welfare.

One factor involves the degree of harm to data collection and analysis. If, for example, an initial Internet offer stands only for a very limited time, and may change based on new data arriving or on the consumer seeking offers elsewhere, then it will be difficult for competitors to match offers. Alternatively, if the unique dataset which affects the conditions of the offer is uni-dimensional (e.g., price), whereas the offer is based on a multi-dimensional calculation, the data-based offer might not enable others to indirectly observe the data.

Another factor involves the externalities and internalities created by the erection of entry barriers in markets that use big data. To illustrate, suppose that there is vertical integration with respect to data collectors and data analysts. On the one hand, such integration might create high (two-level) entry barriers into data markets which, in turn, might create market power and could lead to foreclosing behavior. On the other hand, the vertical integration might overcome free riding and it could increase synergies in the collection and use of data. For example, once the actions of a user indicate his potential interest in a certain product, it may be efficient to immediately provide him with offers about that product, on the platform that he is currently using. Yet, the non-rivalrous characteristic of data, combined with the increased use of algorithms to determine trade conditions and the speed of Internet connection, might reduce the comparative advantage of vertical integration if competitors can speedily copy the on-line offers made by their rivals.

This, in turn, might lead to the erection of barriers to reach the information regarding the first offer by the vertically integrated firm, so that its first offer could not be easily integrated into its rivals' price algorithm. Alternatively, if the integrated firm also operates an on-line platform or a search engine, it might erect barriers for users to reach the trade offers of low-cost suppliers, for example by placing their results lower in the search algorithm's results. These considerations should be taken into account when evaluating the welfare effects of vertical integration.

Finally, it is important to emphasize that, as our analysis shows, the non-rivalrous nature of big data does not imply that its collection does not exhibit entry barriers. Rather, it is these very barriers which create an incentive for firms to compete over the provision of products or services from which they can get access to such information, sometimes even providing them free of charge.[116] These products or services, in turn, may increase social welfare if the added value of the widgets to consumers (in lower prices or higher quality) plus the positive externalities big data creates (e.g., learn to cure diseases faster), exceeds the value of the data to the consumer, plus the potential harm that the gathering of big data might create (e.g., discrimination against certain social groups which results from correlations in the data set).

3. Data as an input

The fact that big data usually serves as an input into other product or service markets, rather than a product of its own, leads to four observations, which have broad relevance. First, the analysis of entry barriers should often extend beyond the specific market under scrutiny to related parts of the data value chain. To illustrate, consider the use of free goods and services in on-line markets. The availability of free goods creates a two-level barrier to entry into data collection. Suppose that Firm A has a comparative advantage in market A which provides data access points. Firm B wishes to compete in market for the data. Firm B faces two entry

[116] Evans and Schmalensee, *supra* note 64.

options: (a) contract with firm A for access to its data; or (b) incur the high costs of entering market A. This implies that some firms will not enter market B, even if they can supply a more efficient product than is currently supplied. This entry obstacle might also lead to a situation in which market B will be monopolized. At the same time, consumers may enjoy lower-priced and higher quality products that are intended to "lure them" to use particular on-line services. An analysis which is based on the realization of barriers in related vertical markets is essential if an analysis of competition and welfare is to be accurate and effective. It is also relevant to market definition.[117]

This observation has not, as of yet, been recognized in all relevant regulatory decisions. For example, in analyzing the advertising-based media, both the FTC and the EU focused on advertising markets and disregarded the dynamics of the markets in which these ads were placed, on the grounds that there is no trade relationship between the online user and the content provider.[118]Such an analysis is, in our view, flawed, because it disregards the competitive effects of the free on-line services market which affect the market for the collection of consumer data, and which in turn affect the dynamics in the advertising market.

The second observation is that a comparative advantage unrelated to data might help overcome entry barriers in big data markets. Tinder serves as a case in point. Its innovation in the on-line dating sites was based on a simple change in how to use the site (left swipe versus right swipe) which catered to consumer needs, rather than on a comparative advantage based on big data.[119]Other examples abound, including the recent entry of Qwant into the search market, which gains its comparative advantage from offering a less data-voracious search service. Indeed, since big data almost never serves as the final product (exceptions might occur where information, by itself, is sought), its effect on competition for the final product must always be determined by a holistic analysis focused on the relative weight of different factors that affect consumer choice.[120] As noted

[117]See James Ratliff and Daniel L. Rubinfeld, *Is There a Market for Organic Search Engine Results and Can Their Manipulation Give Rise to Antitrust Liability*, JOURNAL OF COMPETITION LAW AND ECONOMICS 1 (2014) for a discussion of definition of relevant data markets. Tucker and Wellford, *supra,* note 15, have a similar view. They suggest, however, that personal data could constitute a relevant market if the data were sold to customers. Newman agrees with Ratliff and Rubinfeld that search advertising is a relevant market even when the user data is an essential input. Nathan Newman, *Search, Antitrust and the Economics of the Control of User Data*, YALE JOURNAL ON REGULATION 401 (2014).

[118] European Commission Decision, NewsCorp/Telepiù, Case COMP/M.2876 [2004] OJ L 110/73, paragraph 24 and Competitive Impact Statement at 11-16;United States v. Bain Capital, LLC, 2008 WL 4000820 (D.D.C. 2008) (No. 08-0245).

[119]*See, e.g.*, Sokol and Comerford, *supra* note 18.

[120]*See, e.g.*, Lerner, *supra*, note 17, for such a conclusion with regard to search markets.

above, this fact can explain competition and innovation in many markets that use big data as an input, including the replacement by MySpace by Facebook and Google's successful entry into the search market. At the same time, where big data serves as a product of its own (mostly as an input into other markets), this has only indirect effects on the analysis of entry barriers and market power.

Furthermore, the fact that entry occurs into markets in which big data serves as an input or as an output, does not necessarily lead to the conclusion that low entry barriers also exist in the relevant big data markets. This seemingly simple differentiation appears to be disregarded by some scholars. At the same time, it may, however, signal that even if high barriers exist, their welfare effects on the final consumer might not be high.

Finally, any analysis of the role of big data must also relate to new technological developments, most importantly to the move from traditional vertically-structured industries to networks.[121] In the latter, big data which is collected from multiple sources, including the internet-of-things, serves an input into automated machines, which are digitally connected, to automatically produce or supply the consumer. An automated car serves an example: information about weather, traffic, low-cost gas stations, road difficulties, and the routes requested by other potential users which might wish to share the car, may all automatically affect the route of the car, without requiring any decision by the user. As a result, firms that control all or large part of these networks might enjoy significant comparative advantages in smart products or smart supply chains. This, in turn, might change the industrial landscape as we know it. Where big data serve as an essential input into such networks, the analysis of the effect of entry barriers into its collection, storage, analysis and usage must also relate to their effects on the larger picture.

4. Collection as a by-product

As elaborated above, another characteristic of many types of data is that its collection may be a by-product of other activities. This, in turn, might create a two-level entry problem that may erect high entry barriers in the data-collection market.[122]

Moreover, the fact that each level of the data value chain is potentially characterized by an additional set of entry barriers may affect incentives to

[121] *See* Drexl, *supra* note 78.

[122] The "classic" two-level entry issue arose in *U.S. v. Microsoft*, 253 F. 3d 34 (2001), where the Court recognized an "applications barrier to entry" whereby to compete in an operating systems market one also needed one or more widely utilized applications, such as an office suite.

share such data. For example, if entry barriers in data storage and analysis are high, firms whose business model is not necessarily based on big data might have stronger incentives to sell the raw data they collect, thereby limiting the two-level entry barrier. Yet it is important to emphasize that the fact that the data would have been harvested in any case, or that its collector has no other use for it, does not necessarily affect its market value or price.

5. Big Data may exhibit Strong Network Effects

As noted previously, some types of big data may exhibit strong network effects. These include network effects arising from the fact that the scale, scope or speed of data collection may positively affect the accuracy of information that can be discerned from it. Multi-sidedness of a market strengthens such network effects.

These entry barriers have led the OECD to observe that big data favors market concentration and dominance," and that "data driven-markets can lead to a "winner takes all" result where concentration is a likely outcome of market success."[123] As our research has shown, this is not true of all data-driven markets. Much depends on the height of entry barriers which characterize the specific market. It also depends on market structure: whether several services are provided together through an intermediary, or whether each service is provided by a stand-alone provider. Yet when network effects are substantial, and especially when they are coupled with other entry barriers, they can generate significant competitive effects. This is especially so when first-mover data-based comparative advantages are high or when some firms are poised to become super-platforms of data and services.[124]

6. Data might strengthen discrimination

A final characteristic of some types of big data which may affect the competitive analysis centers on the ability to price discriminate more easily among down-stream consumers, should the big data provide information regarding consumer preferences. Usually, price discrimination affects only immediate consumers. With big data, this implies that the data collector or analyzer often has the ability to demand different prices from those buying the data, depending on the buyers' elasticities of demand. Yet the more important price discrimination effects are often created at a lower level of the supply chain. In big-data driven markets, in which the data

[123] OECD, Data-Driven Innovation, *supra* note 11, p. 7.

[124] *See also* Stucke and Grunes, Big Data, *supra* note 22, p. 251-54, 275.

reveal consumer preferences, the ability to price discriminate may be substantial. The question then becomes whether such price discrimination increases overall welfare, or simply (or mostly) benefits the user of the data.[125]

A related question is how, if at all, the competitive conditions in big data markets affect price discrimination in the consumer-goods market. Much depends on the structure of the market exploiting such data. A monopolistic user of data may engage in first-degree price discrimination. Should such discrimination be based on the data on consumers' preferences rather than on the relative quality of the product purchased, this could reduce consumer welfare substantially given the ability of the monopolist controlling to data to effectively extract consumer surplus.

Introducing competition among the users of such data might not, however, necessarily increase welfare. Much depends on consumer conduct and the ultimate competitive equilibrium. Assume, for example, that consumers do not have much knowledge about other suppliers, or that they exhibit behavioral traits and biases, such as accepting the first offer they receive while not spending time comparing it with other offers, or exploring only the first links in their search query for comparable products. Such conduct might be rational, especially with regard to low-priced products. Under scenarios such as this, the fact that multiple data users have access to similar data would not necessarily reduce the price discrimination problem. Rather, it would allow all owners of datasets to engage in discrimination. If, however, consumers engage in substantial (albeit costly) searches, some suppliers might respond by offering better trade conditions, which would reduce the use and effects of price discrimination. Algorithms that search for better offers for consumers might serve to reduce this problem. The welfare effects of the erection of barriers for such searches should be analyzed.

Observe further that entry barriers into big data markets do not necessarily affect consumer welfare. This is because such welfare effects depend, other things equal, on the effect of the data-based-information on the price and quality of the final product or service. Assume, for example, that information on consumer preferences significantly increases the ability of one competitor to market his products effectively by targeting potentially interested consumers. Other suppliers of competing goods, which lack such targeted marketing information, may nonetheless enjoy other comparative advantages such as patents, location, unique human and

[125] For an overview of the economics of price discrimination, see ROBERT S. PINDYCK AND DANIEL L. RUBINFELD, MICROECONOMICS, (8th ed., 2013), Chapter 11. See also, DENNIS W. CARLTON AND JEFFREY M. PERLOFF, MODERN INDUSTRIAL ORGANIZATION (3rd ed., 2000), Chapter 9.

physical resources, and reputation. In other situations big data advantages may increase the incentives of other firms to compete not only on the big-data-based-information, but on other dimensions of the product, including quality and price. The higher the market reward for informational accuracy, the larger the advantage of big data that must be overcome by other advantages. Where this is true, big data should be treated no differently than other inputs that create comparative advantages.[126] Further observe that, as noted above, big data might affect wider aspects of social welfare, such as equal opportunity.

7. Additional observations

Due to its above features, often a multi-faceted balance needs to be struck between anti-competitive effects and pro-competitive and public good justifications (e.g., synergies, motivations for innovation, privacy concerns) with regard to the erection or lowering of entry barriers. An important part of the welfare analysis focuses on the uses of the data and how such uses affect welfare. Big data on how best to treat an illness used by doctors is not the same as big data on betting habits used by firms that offer gaming opportunities.

It is also noteworthy that big data might change the competitive dynamics across many markets. One example involves the market for advertising. Traditional ways to reach potential consumers included banners and newspaper advertising. More accurate information on consumers' characteristics, arising from big data that either relates to the specific consumer or to groups the consumer belongs to, allow for more targeted advertising, often through Internet sites that reach each consumer specifically. This, in turn, also affects the delineation of relevant markets.

8. Adding an International Dimension

To this point, we have largely disregarded the international dimension: how the data's characteristics affect the conduct of, and competition between, international firms. While this is a subject that justifies a paper of its own, several observations are in order. Entry barriers into big data markets may differ from one jurisdiction to another, thereby creating higher obstacles to the collection, storage, analysis or usage of big data in certain jurisdictions. However, the characteristics of the data may allow for cross-border spillovers in data analysis. For example, where consumers across

[126] For a similar conclusion see Balto and Lane, *supra* note 104, at 3-4.

certain jurisdictions are likely to be relatively similar, data collected with regard to consumers in one jurisdiction can be used in other jurisdictions, thereby overcoming barriers to the collection and analysis of such data in the latter. A large international scale might also make it easier for firms to enjoy scale and scope economies in data collection and analysis, and for consumers to enjoy stronger network effects. This, in turn, creates a comparative advantage to firms operating in multiple jurisdictions. It might also make it more difficult for national firms to compete in their own jurisdictions. Yet firms do not necessarily need to operate in other jurisdictions in order to use some of their advantages to overcome entry barriers. Storage provides a good example. Due to the data's transferability, storage can be performed in jurisdictions with better storage capabilities, while used in others. Such cross-border effects must be taken into account when analyzing the height of entry barriers.

Another issue arising from the internationalization of big data involves regulation. In today's world, regulation is mainly based on local welfare considerations (mostly country-specific and in some instances region-specific). Moreover, such regulation might not necessarily be based on competition considerations, but rather on wider ones, including industrial policy and human rights. An interesting set of questions arises whether differences in legal barriers may lead to a race to the bottom, whether limitations in certain jurisdictions might create significant externalities that would affect welfare elsewhere, and whether some type of harmonized global regulation might be more efficient.

To sum up, big data creates new and sometimes complex issues regarding competition and social welfare. The implications are widespread, ranging from data-motivated acquisitions, to the erection of artificial barriers to markets in the data-value chain. The challenge is to find the optimal balance between the clashing effects on social welfare analyzed above.

While beyond the scope of this paper, we have on occasion noted implications for regulatory and competition policy. Any analysis which groups together all big data markets in a broad-brush analysis, or assumes that all big data markets are open to competition because of the non-rivalrous nature of data, is likely to be problematic. To exemplify, a recent OECD Report suggested that the economics of big data "[favors] market concentration and dominance."[127] Yet unless one recognizes the unique entry barriers into each relevant market under scrutiny, such suggestions

[127] OECD Report, supra note 11, at 7.

are not helpful.[128] To take another example, in “Big Mistakes Regarding Big Data,” Tucker and Wellford claim that “Relevant data are widely available and often free,” and therefore antitrust has a limited role to play.[129] To the contrary, we have shown that this assumption is not necessarily true, once one recognizes entry barriers down the data value chain.

Finally, our article has shown that antitrust may well be relevant with respect to some aspects of big data markets. While much depends on case-specific facts, some features of big data markets create a solid basis for theories of harm to competition and welfare that should not be ignored. The U.S. merger of *Baazarvoice/Power-Reviews* serves as such an example: there the Department of Justice found that the data created an entry barrier into the market for rating and review platform.[130] Other cases may well follow, especially where the antitrust authorities exhibit the necessary flexibility to incorporate the unique features of big data markets into their analysis.

CONCLUSIONS

Big data has become a most valuable resource in our digital world. Data analysis has developed from a tool to expand knowledge and efficiency to an actual commodity. The collection and analysis of big data has undoubtedly increased social welfare. However, big data markets are also often characterized by entry barriers which, in turn, have the potential to create durable market power in data-related markets or to serve as a basis for anti-competitive conduct.

This paper explored the entry barriers into markets in the data value chain and analyzed some of their implications on the competitive analysis of such markets. We hope we have advanced the study of the potential

[128] Eur. Data Prot. Supervisor, REPORT OF WORKSHOP ON PRIVACY, CONSUMERS, COMPETITION AND BIG DATA 1 (2014), https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/BW/2014/14-07-11_EDPS_Report_Workshop_Big_data_EN.pdf. (“the number of “big data” related mergers and acquisitions more than doubled between 2008 and 2012”).

[129] Tucker and Wellford, *supra* note 15, at 1.

[130] DOJ, Antitrust Division, Competitive Impact Statement, 13-cv-00133 WHO, dated 08.05.2014, http://www.justice.gov/atr/case-document/file/488826/download, at 5. The French Competition Authority has recently ordered a gas supplier to grant its competitors access to some of the dataitcollectedasaproviderofregulatedoffers,in order to allow all suppliers to have the same level of relevant information to make offers to consumers. French Competition Authority, Decision 14-MC-02 of 09.09.2014.*See also*, the French and German Report, *supra* note 14.

competitive effects of big data, in a way which could better enable regulators and scholars to create or suggest improvements in the law.