

# Big Mistakes Regarding Big Data

**Darren S. Tucker and Hill B. Wellford**

The ability to store and analyze enormous quantities of data—so-called big data—offers tremendous benefits to both individuals and businesses. Big data analysis has helped increase economic output, reduce crime, improve public health and safety, increase voter turnout, boost energy efficiency, improve weather forecasts, and enhance agricultural yields. The use of big data has become widespread. Once the province of big businesses and governments, analysis of large datasets has become so inexpensive as to be an option even for the prototypical garage-based start-up company. Relevant data are widely available and often free.

Recently, there have been calls for antitrust intervention to address concerns related to big data, particularly big data consisting of personal information.<sup>1</sup> Some privacy regulators, policy organizations, and businesses claim that big data presents a significant and durable entry barrier for online services that has led to entrenchment of large firms. According to these proponents of heightened antitrust scrutiny, large online firms should face antitrust liability for refusing to provide user data in their possession to rivals or for collecting additional user data by expanding into new product lines (whether through acquisition or organic growth). The appropriate remedy for these alleged antitrust violations, they say, is to ensure that incumbents do not have advantages over entrants in terms of access to consumer data. Some also have urged the competition agencies to expand the concept of a relevant antitrust market to include data markets, even when that data is not marketed to customers. Implicit in many of these claims is the belief that a few large companies control most big data or have significant market power due to their possession of vast data collections.

Contrary to the views of these advocates for antitrust enforcement, the acquisition and use of big data by online firms is not the type of conduct captured by the antitrust laws. Online markets are notable for their low entry barriers and typically do not require big data for entry. Instead, most online service providers, including social media, search, and retail, use big data to improve their services once a customer base is established. As a result, the collection and analysis of big data have enhanced competition and improved product offerings. In addition, the notion of a relevant

■ **Darren Tucker and Hill Wellford** are partners in the Washington, DC office of Morgan, Lewis & Bockius LLP. Darren is Editorial Chair of *The Antitrust Source*. The authors represent Google on certain matters relating to the subject of this article but did not receive funding for writing this article and the views expressed in this article are those of the authors alone.

<sup>1</sup> See, e.g., EUR. DATA PROT. SUPERVISOR, **PRIVACY AND COMPETITIVENESS IN THE AGE OF BIG DATA: THE INTERPLAY BETWEEN DATA PROTECTION, COMPETITION LAW AND CONSUMER PROTECTION IN THE DIGITAL ECONOMY** (preliminary opinion Mar. 2014) [hereinafter EDPS REPORT]; Nathan Newman, *Search, Antitrust and the Economics of the Control of User Data*, 31 YALE J. ON REG. 401 (2014); Letter from Jeff Chester, Exec. Dir., Ctr. for Digital Democracy & Edward Mierzwinski, Dir., U.S. PIRG Educ. Fund, to Edith Ramirez, Chairwoman, Fed. Trade Comm'n (Nov. 5, 2014); FairSearch, *Nest Acquisition: Another Brick in Google's Great Wall of Data* (Jan. 22, 2014). In addition, in its November 27, 2014 resolution calling on the European "Commission to enforce EU competition rules decisively" with respect to online search, the European Parliament cited "the possibility [search engines] have of commercialising secondary exploitation of information obtained" as one basis for its concerns. **Resolution on Supporting Consumer Rights in the Digital Single Market**, EUR. PARL. DOC. 2014/2973(RSP) (2014).

market consisting of internally used data is inconsistent with longstanding precedent that recognizes a market only where a product or service is sold to consumers. Furthermore, remedies that have been proposed to limit incumbents' collection or use of big data or to require forced sharing with rivals are likely to harm competition and raise privacy concerns.

### What Is Big Data?

Generally, "big data" refers to a collection of data sets so large and complex that traditional database systems cannot effectively manage or process the information.<sup>2</sup> Information technology research company Gartner defines big data as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."<sup>3</sup> As this widely-adopted "3V" definition suggests, big data is not defined only by volume, but also by the complexity of the data (e.g., different types of structured or unstructured data, including text, image, audio, and video files) and the need for the data to be collected and analyzed rapidly.<sup>4</sup>

*There are significant  
benefits to consumers,  
businesses, and  
government agencies  
from collection and  
analysis of big data.*

There are significant benefits to consumers, businesses, and government agencies from collection and analysis of big data. Companies have been able to offer entirely new products and services (e.g., real-time traffic information), enhance existing products and services (e.g., personalized music or video recommendations), and better market their products to consumers (e.g., targeted promotions). Targeted advertising subsidizes free services and content on the Internet, as well as many free mobile apps, and has helped fund the meteoric growth of and innovation on the Internet.

Big data is everywhere. As a 2014 White House report noted, "We live in a world of near-ubiquitous data collection."<sup>5</sup> The sources of data continue to grow, with more recent examples including networked sensors, "wearables," smart appliances, and geospatial technologies. The number of firms with extensive data collection and processing capabilities is vast, including online and offline retailers, grocery stores, online advertising networks, search engines, social networking sites, Internet service providers (ISPs) and cable companies, financial institutions, insurance companies, and data brokers. Nearly all online providers collect data from their users; one online privacy company claims that the typical Internet user is tracked over 100 times during each browsing session.<sup>6</sup>

Big data is used by organizations of all sizes. Small businesses, entrepreneurs, and government agencies, in addition to large companies, are avid users of big data.<sup>7</sup> As FTC Chairwoman

<sup>2</sup> See JAMES MANYIKA ET AL., MCKINSEY GLOBAL INST., **BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY** 1 (June 2011) [hereinafter MCKINSEY REPORT] ("Big data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.").

<sup>3</sup> **Big Data Definition**, IT GLOSSARY, GARTNER. Some add a fourth V for "veracity."

<sup>4</sup> EXECUTIVE OFFICE OF THE PRESIDENT, **BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES** 2 (2014) [hereinafter WHITE HOUSE REPORT] ("Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data.").

<sup>5</sup> *Id.* at 4; see also MCKINSEY REPORT, *supra* note 2, at 2 ("Digital data is now everywhere—in every sector, in every economy, in every organization and user of digital technology.").

<sup>6</sup> **DoNotTrackMe: How It Works**, ABINE (last visited Dec. 7, 2014).

<sup>7</sup> WHITE HOUSE REPORT, *supra* note 4, at 5–6, 22–38 (governments), 39 (small firms), 48 (entrepreneurs); MCKINSEY REPORT, *supra* note 2, at 106 (individuals and SMEs); FED. TRADE COMM'N, **DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY** 47–48 (2014) [hereinafter FTC DATA BROKER REPORT] (observing that data brokers facilitate small businesses reaching consumers with innovative products, leading to "increased competition").

Edith Ramirez observed, “[B]ig data is no longer the province of a few giant companies.”<sup>8</sup> As a result, innovative uses of big data by smaller firms have the potential to introduce new forms of competition and disrupt established industries.<sup>9</sup>

The cost of collecting big data is very low and continues to decline.<sup>10</sup> Most companies generate a tremendous volume of “exhaust data,” i.e., data created as a by-product of their usual activities, such as sales transactions and interactions with customers.<sup>11</sup> The cost of collecting this data is virtually zero. Storing and analyzing data are also inexpensive,<sup>12</sup> and the ability to derive useful insights from big datasets continues to improve with the development of increasingly sophisticated software. There is even widely used, free, open-source software available to analyze (Hadoop) or host (Cassandra, HBase) big datasets.

Big data is widely available for purchase and is inexpensive. There is a “near-continuous collection, transfer, and re-purposing of information in a big data world.”<sup>13</sup> Consumer data and other information can be purchased from an array of companies, including data brokers. One data broker alone has “multi-sourced insight into approximately 700 million consumers worldwide.”<sup>14</sup> These profiles can contain “thousands of pieces of data” on an individual.<sup>15</sup> Basic demographic information sells for about \$0.0005 (5 one-hundredths of a cent) per person; even a detailed profile of an individual about to make a purchase—and therefore very valuable to advertisers—typically costs well under one dollar.<sup>16</sup>

Big data is non-rivalrous. In other words, collecting a particular piece of data does not prevent other companies from collecting identical data by similar or other means. Using multiple providers for the same service (user multi-homing) and the common practice of website operators using multiple ad networks and analytic firms make it easier for multiple providers to collect relevant user

---

<sup>8</sup> Edith Ramirez, Chairwoman, Fed. Trade Comm’n, *The Privacy Challenges of Big Data: A View from the Lifeguard’s Chair* 3 (Aug. 19, 2013). See also Edd Dumbill, *Volume, Velocity, Variety: What You Need to Know About Big Data*, FORBES, Jan. 19, 2012 (“Big data processing is eminently feasible for even the small garage startups, who can cheaply rent server time in the cloud.”).

<sup>9</sup> MCKINSEY REPORT, *supra* note 2, at 6 (“[E]stablished competitors and new entrants alike will leverage data-driven strategies to innovate, compete, and capture value. Indeed, we found early examples of such use of data in every sector we examined.”).

<sup>10</sup> WHITE HOUSE REPORT, *supra* note 4, at 1 (explaining that the growth of big data is fueled by “the cratering costs of computation and storage”); Ramirez, *supra* note 8, at 3 (observing that the “phenomenal growth in storage and analytic power” has been accompanied by an equally dramatic drop in costs).

<sup>11</sup> MCKINSEY REPORT, *supra* note 2, at 1; Edith Ramirez, Chairwoman, Fed. Trade Comm’n, *Big Data: A Tool for Inclusion or Exclusion?* 3 (Sept. 15, 2014) (“The proliferation of connected devices, the plummeting cost of collecting, storing, and processing information, and the ability of data brokers and others to combine offline and online data mean that companies can accumulate virtually unlimited amounts of consumer information and store it indefinitely.”).

<sup>12</sup> MCKINSEY REPORT, *supra* note 2, at 2 (“The ability to store, aggregate, and combine data and then use the results to perform deep analyses has become ever more accessible as trends such as Moore’s Law in computing, its equivalent in digital storage, and cloud computing continue to lower costs and other technology barriers.”).

<sup>13</sup> WHITE HOUSE REPORT, *supra* note 4, at 39. See also *id.* at 50 (“[Big data] is bought, bartered, traded, and sold. An entire industry now exists to commoditize the conclusions drawn from that data.”).

<sup>14</sup> STAFF OF S. COMM. ON COMMERCE, SCI., AND TRANSP., OFFICE OF OVERSIGHT & INVESTIGATIONS, *A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE, AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES* 12 [hereinafter SENATE COMMITTEE REPORT] (quoting Acxiom Corp., Annual Report (Form 10-K) (May 29, 2013)).

<sup>15</sup> WHITE HOUSE REPORT, *supra* note 4, at 44. See also FTC DATA BROKER REPORT, *supra* note 7, at 47 (“[O]ne of the nine data brokers has 3000 data segments for nearly every U.S. consumer.”); SENATE COMMITTEE REPORT, *supra* note 14, at 12 (“Some of the companies maintain thousands of data points on individual consumers, with one providing the Committee a list of approximately 75,000 individual data elements that are in its system.”).

<sup>16</sup> Emily Steel, *Financial Worth of Data Comes in at Under a Penny a Piece*, FIN. TIMES, July 12, 2013.

data. For example, if one ad network determined that the user of a particular mobile device lived in Connecticut, liked to travel, and owned a dog, there is nothing to prevent another ad network from learning the same information—indeed, for a frequent Internet user, it is likely that dozens of firms will create a similar profile. Redundant data are so common as to cause problems for data brokers.<sup>17</sup>

Big data can become stale quickly. Historical data can be analyzed for trends but has comparatively little value when used for real-time decisions, such as which ad to serve. As one study found, “90% of the data in the world today has been created in the last two years. . . . 70% of unstructured data is stale after only 90 days.”<sup>18</sup> Data collection and analysis is increasingly occurring on a real-time or near-real-time basis.<sup>19</sup>

These characteristics—ubiquity, low cost, wide availability, and fleeting value—make big data different from the industry structures typically seen as conducive to competition problems. That, of course, does not preclude a violation of the antitrust laws due to conduct involving big data, but it does suggest that courts and regulators should proceed cautiously when presented with claims that big data is the source of a competition problem.

### Can Data Be Its Own Product Market?

Some privacy regulators and other writers have asserted that personal information collected in connection with the provision of online services should be considered an intangible asset for antitrust purposes and have called for markets to be defined around personal data, even when that information is not marketed to customers.<sup>20</sup> Such a view lacks support under U.S. and EU competition law.<sup>21</sup> Personal data used as an input to another product cannot constitute a relevant product market because a product market presupposes that a product or service is available to customers. Only where data is sold to customers could providing that information potentially constitute a relevant market.

In both the United States and Europe, market definition turns primarily on demand substitution, i.e., the degree to which customers will substitute one product for another. This is because relative to other considerations, such as supply substitutability and potential competition, demand substitution offers the most immediate and effective disciplinary force on suppliers of a given product, particularly in relation to their pricing decisions. The FTC, DOJ, and EC apply the “hypothetical monopolist” test, which asks whether customers would switch to substitutes in response

<sup>17</sup> See generally FTC DATA BROKER REPORT, *supra* note 7, at 14 (“[E]ach data broker utilizes multiple sources for similar data. For example, one of the data brokers in this study obtains consumers’ contact information from twenty different sources.”).

<sup>18</sup> RIS, *ANALYTICS INSIGHTS DELIVER COMPETITIVE DIFFERENTIATION 2* (July 2013) (quoting Citi Research 2013 Retail Technology Deep Dive). See also *Big Data, for Better or Worse: 90% of World’s Data Generated over Last Two Years*, SCIENCEAILY (May 22, 2013) (“A full 90% of all the data in the world has been generated over the last two years.”).

<sup>19</sup> WHITE HOUSE REPORT, *supra* note 4, at 5. As one author aptly illustrated, an insurer determining that a claim is fraudulent three months after paying out the claim offers little value compared to identifying the fraud while processing the claim. See Nick Millman, *The Need for Speed with Big Data*, BLOOMBERG BUSINESSWEEK, Aug. 6, 2013.

<sup>20</sup> EDPS REPORT, *supra* note 1, § 4.1; Pamela Jones Harbour & Tara Isa Koslov, *Section 2 in a Web 2.0 World: An Expanded Vision of Relevant Product Markets*, 76 ANTITRUST L.J. 769, 773 (2010) (“[W]e suggest the definition of markets for data, separate and apart from markets for the services fueled by these data.”).

<sup>21</sup> One proponent of this approach acknowledges that “[d]efining a market for user data may be unusual under traditional market definition principles.” Pamela Jones Harbour, *Competition & Privacy in Markets of Data*, Comments to European Parliament 5 (Nov. 26, 2012).

*If the courts or antitrust*

*agencies were to define*

*relevant markets*

*around inputs like*

*consumer data,*

*antitrust analysis would*

*become more complex,*

*less accurate, and less*

*predictable.*

to a hypothetical small relative price increase in the products and areas under investigation.<sup>22</sup> The U.S. federal courts often apply the *Brown Shoe* test, under which “[t]he outer boundaries of a product market are determined by the reasonable interchangeability of use or the cross-elasticity of demand between the product itself and substitutes for it.”<sup>23</sup> Thus, under accepted market definition tests, prerequisites for defining a relevant market include the sale of a product or service to customers and a means of determining substitution among products by customers.

Personal information collected by a producer but not sold to customers cannot satisfy the hypothetical monopolist test or the *Brown Shoe* test: there is no sale, no customers, and no product substitution. To our knowledge, none of the FTC, DOJ, EC, General Court, European Court of Justice, or U.S. federal courts has ever defined a separate market for internally generated and used (i.e., captive) personal data.

In the absence of sales, the rationale for defining a relevant market consisting of personal information is unclear. Why define a product market if there is neither a “product” nor a “market”? The principal goal of market definition is to draw a line between products that substantially compete and those that do not.<sup>24</sup> By definition, there can be no competition where there are and will be no sales.

If the courts or antitrust agencies were to define relevant markets around inputs like consumer data, antitrust analysis would become more complex, less accurate, and less predictable. There would be almost no end to the number of relevant markets that would require analysis in most mergers.

The FTC’s review of Nielsen’s acquisition of Arbitron illustrates how markets should be defined where personal data are not for sale—that is, as a key input but not as a separate relevant market. In that case, the FTC found that the merging companies had “the most accurate and preferred sources of individual-level demographic data for [television and radio] audience measurement purposes.”<sup>25</sup> This information was alleged to be a critical requirement for the development of a service to measure audiences across multiple media platforms. The FTC did not define a market around the audience data or find a competitive issue related to the data, but instead concluded that the competitive concern was in the downstream market for national syndicated cross-platform audience measurement services. The agency observed that the demographic data were used internally by the two firms and were not sold to third parties.

<sup>22</sup> U.S. Dep’t of Justice & Fed. Trade Comm’n, Horizontal Merger Guidelines § 4.1.1 (2010) [hereinafter U.S. Merger Guidelines]; Commission Notice on the Definition of Relevant Market for the Purposes of Community Competition Law, 1997 O.J. (C 372) 5, ¶ 17 [hereinafter EC Market Definition Guidelines].

<sup>23</sup> United States v. Brown Shoe Co., 370 U.S. 294, 325 (1962). See also Eastman Kodak Co. v. Image Technical Servs., 504 U.S. 451, 469 (1992) (defining the relevant market by reference to “the extent to which consumers will change their consumption of one product in response to a price change in another”); United States v. E.I. du Pont de Nemours & Co., 351 U.S. 377, 393 (1956) (explaining that the relevant market consists of “commodities reasonably interchangeable by consumers for the same purposes”); U.S. Merger Guidelines, *supra* note 22, § 4 (“Market definition focuses solely on demand substitution factors, i.e., on customers’ ability and willingness to substitute away from one product to another . . .”); EC Market Definition Guidelines, *supra* note 22, ¶ 7 (“A relevant product market comprises all those products and/or services which are regarded as interchangeable or substitutable by the consumer, by reason of the products’ characteristics, their prices and their intended use.”).

<sup>24</sup> Tampa Elec. Co. v. Nashville Coal Co., 365 U.S. 320, 328 (1961) (explaining that the relevant market is “the area of effective competition”); Geneva Pharm. Tech. Corp. v. Barr Labs., Inc., 386 F.3d 485, 496 (2d Cir. 2004) (“The goal in defining the relevant market is to identify the market participants and competitive pressures that restrain an individual firm’s ability to raise prices or restrict output.”); Thurman Indus., Inc. v. Pay ‘N Pak Stores, Inc., 875 F.2d 1369, 1374 (9th Cir. 1989) (“For antitrust purposes, defining the product market involves identification of the field of competition: the group or groups of sellers or producers who have actual or potential ability to deprive each other of significant levels of business.”).

<sup>25</sup> 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).

The European Commission's review of the Facebook/WhatsApp transaction is another example of properly defined markets where data are not for sale. Despite concerns raised by some third parties that the transaction would bolster Facebook's access to data, the Commission declined to define a market for data because the parties did not sell any of the data they collected to advertisers or other third parties. The Commission explained that it "has not investigated any possible market definition with respect to the provision of data or data analytics services, since . . . neither of the Parties is currently active in any such potential markets."<sup>26</sup>

In contrast, personal data could constitute a relevant market if sold to customers. Although not involving personal data, examples of recent FTC and DOJ cases with data-related relevant markets include CoreLogic/DataQuick (assessor and recorder data),<sup>27</sup> Dun & Bradstreet/QED (educational marketing data),<sup>28</sup> and Thomson/Reuters (financial information).<sup>29</sup> The U.S. courts have done so as well: in CCC/Mitchell, the district court defined the relevant market as automobile damage estimating software, a key part of which was a proprietary parts and labor database.<sup>30</sup> The EC also has reviewed several mergers between firms providing data or data-related services, including Thomson/Reuters,<sup>31</sup> Oracle/Sun (database software systems),<sup>32</sup> WPP/TNS (TAM services to measure media audiences),<sup>33</sup> Telefónica UK/Vodafone UK/Everything Everywhere JV (data analytics services for online and offline advertising),<sup>34</sup> and IBM Italia/UBIS (provision of data center services).<sup>35</sup>

Thus, there is no shortage of precedent in the United States and Europe for defining a market around data sold to customers, but there is no legal support in either region for defining a relevant market around data that firms generate and use solely as an input.

### Is Data a Barrier to Entry?

Recently, a number of writers and advocacy groups in the United States and Europe have asserted that big data—in particular, user data—is a critical requirement for offering online services.<sup>36</sup> They argue that extensive user information is needed to sell advertising and improve online serv-

<sup>26</sup> Case COMP/M.7217—Facebook/WhatsApp, Comm'n Decision, ¶ 72 (Mar. 10, 2014).

<sup>27</sup> Complaint ¶ 5, CoreLogic, Inc., FTC Docket No. C-4458 (2014). Darren Tucker represented the divestiture buyer in the CoreLogic case.

<sup>28</sup> Complaint ¶ 11, Dun & Bradstreet Corp., FTC Docket No. 9342 (2010).

<sup>29</sup> Complaint ¶¶ 14, 20, United States v. Thomson Corp., Case No. 1:08-cv-00262 (D.D.C. 2008).

<sup>30</sup> FTC v. CCC Holdings, Inc., 605 F. Supp. 2d 26, 38–39 (D.D.C. 2009). Darren Tucker represented Mitchell in this matter.

<sup>31</sup> Case COMP/M.4726—Thomson Corp./Reuters Grp., Comm'n Decision (Feb. 19, 2008).

<sup>32</sup> Case COMP/M.5529—Oracle/Sun Microsystems, Comm'n Decision (Jan. 21, 2010).

<sup>33</sup> Case COMP/M.5232—WPP/TNS, Comm'n Decision (Sept. 23, 2008).

<sup>34</sup> Case COMP/M.6314—Telefónica UK/Vodafone UK/Everything Everywhere/JV, Comm'n Decision (Sept. 4, 2012).

<sup>35</sup> Case COMP/M.6921—IBM Italia/UBIS, Comm'n Decision (June 19, 2013).

<sup>36</sup> See, e.g., EDPS REPORT, *supra* note 1, ¶¶ 66–68 ("Powerful or dominant undertakings are able to . . . create barriers to entry through their control of huge personal datasets . . . [which] could prevent the development of competing products from competitors."); Newman, *supra* note 1, at 401 ("Google's ever expanding control of user personal data and its critical value to online advertisers creates an insurmountable barrier to entry for new competition."); 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."); Frank Pasquale, *Paradoxes of Digital Antitrust: Why the FTC Failed to Explain Its Inaction on Search Bias*, HARV. J. L. & TECH., July 2013 (Occasional Paper Series) ("Google itself controls the chief input into better search services: the data that engineers need in order to better personalize results."); Damien Geradin & Monika Kuschewsky, *Competition Law and Personal Data: Preliminary Thoughts on a Complex Issue* 2 (Feb. 12, 2013) ("The acquisition of large volumes of data by 'first mover' providers may, however, raise barriers to entry and thus deprive users from the benefits of competition.").



ices, with search and social media the most frequently cited examples. As a result, existing big data providers allegedly have an entrenched position that cannot be disrupted because new entrants are at a severe disadvantage in accessing data.

This view reflects a fundamental misunderstanding of the role of consumer data and other types of data in developing and commercializing online services. The fact that some established online firms collect a large volume of data from their customers or other sources does not mean that new entrants must have the same quantity or type of data in order to enter and compete effectively. To the contrary, little if any consumer data is needed to begin offering most online services, and new entrants are unlikely to be at any significant disadvantage relative to incumbents in terms of data collection or analysis. Relevant data are widely available and inexpensive. In addition, history has shown that entry barriers generally—not just those related to data requirements—are low in the online space, as evidenced by the tremendous amount of entry and rapid gains often enjoyed by innovative new challengers.

*The fact that some established online firms collect a large volume of data from their customers or other sources does not mean that new entrants must have the same quantity or type of data in order to enter and compete effectively.*

As an initial matter, the idea that an input like data could be a barrier to entry because entrants do not have the same amount of that input seems questionable—particularly where, as here, that input is ubiquitous, low cost, and widely available. We would not say, for example, that labor is a barrier to entry because an incumbent has more workers than an entrant, or that a factory is a barrier to entry just because an incumbent has more square footage than an entrant's factory. In other words, lack of asset equivalence should not be a sufficient basis to define a barrier to entry.

An entrant that needs personal data can collect relevant information from its users once the service is operational. Data collected in this manner is free or nearly so. Entering the market and then collecting and analyzing user data is not a theoretical approach but rather the very model followed by many of the leading online firms when they were startups or virtual unknowns, including Google, Facebook, Yelp, Amazon, eBay, Pinterest, and Twitter. For example, Google replaced Yahoo as the leading general search engine within a few years of its entry despite Yahoo having user data on several hundred million users and offering personalized search results prior to Google.<sup>37</sup> Likewise, Facebook rapidly displaced MySpace as the leading social media service, despite the latter company's extensive user data. We are aware of no principled explanation for why this approach is infeasible today.

Entrants can also acquire relevant data from third parties. As previously noted, possession of big data is widespread, and third-party sourcing options, including other online providers or data brokers, continue to expand and fall in cost.<sup>38</sup> For example, search engine DuckDuckGo claims to base its results on over 100 sources, including "crowd-sourced sites (like Wikipedia, which are stored in our own index), Yahoo! (through BOSS), Yandex, Yelp, and Bing."<sup>39</sup> Similarly, Apple's Siri service relies on information provided by Yelp, Yahoo Local, and Rotten Tomatoes, among others.<sup>40</sup> Data brokers source most of their data from other data brokers.<sup>41</sup>

The fact that incumbents may collect a large volume of data says little about the data needs of

<sup>37</sup> See Chris Sherman, *Google My Search History Personalizes the Web*, SEARCH ENGINE WATCH (Apr. 19, 2005) (observing that other search engines had "offered personalized search for some time" at the time Google introduced its personalized search feature).

<sup>38</sup> WHITE HOUSE REPORT, *supra* note 4, at 39 ("[A]ccess to data and the tools for processing it are [becoming] further democratized.").

<sup>39</sup> *Sources*, DUCKDUCKGO (last visited Dec. 7, 2014).

<sup>40</sup> *Siri*, WIKIPEDIA (last visited Dec. 7, 2014).

<sup>41</sup> FTC DATA BROKER REPORT, *supra* note 7, at 46.

new entrants because entrants need not follow the same business model as incumbents. In fact, new entrants in established online product areas tend to focus on particular functionality, customer segments, or user interests. For example, there has been a proliferation of dating sites focused on particular religions, ethnicities, or locations, rather than replicating the across-the-board approaches of Match and eHarmony. Likewise, in social networking, new entrants have tended to focus on particular types of content (Instagram, Twitter, Pinterest) or interests (Reddit, LinkedIn), rather than offering a broader Facebook-like approach. New search engines have differentiated themselves from Google, Bing, and Yahoo by focusing on particular features (DuckDuckGo, Blekko) or verticals (shopping, local information, travel). As a result, the data requirements of new online competitors, if they exist at all, will frequently be far more modest, and qualitatively different, than for established firms. And to the extent that new entrants eventually expand into the space of the larger, more broadly focused firms, they are likely to do so gradually, in which case any data they need can be self-generated.

The antitrust agencies and courts, to our knowledge, have never found user data to be a barrier to entry for online services. In its Google/DoubleClick investigation, the FTC concluded that incumbents' user information is not a barrier to entry for online advertising. The agency explained that "neither the data available to Google, nor the data available to DoubleClick, constitutes an essential input to a successful online advertising product."<sup>42</sup> Likewise, in connection with closing its investigation into Facebook's acquisition of WhatsApp, the European Commission explained that "there are currently a significant number of market participants that collect user data alongside Facebook," including Google, Apple, Amazon, eBay, Microsoft, AOL, Yahoo, Twitter, IAC, LinkedIn, Adobe and Yelp.<sup>43</sup> In addition, "there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook's exclusive control."<sup>44</sup>

Competition agencies and courts have, to be sure, concluded that data-related entry barriers may exist for the sale of data that cannot be sourced from consumers or big data marketplaces. In the FTC, DOJ, and EC cases identified in the prior section, there were sizeable costs to collect the data at issue. Notably however, none of those cases involved data collected from consumers over the Internet.

The fact that there has been—and continues to be—so much entry in online services suggests that entry barriers, in general, are low for these areas and that users are not locked into current providers. Even in more "mature" online categories where incumbents may have significant stores of consumer data, there are a large number of suppliers and entry continues apace:

- **Social Media:** The number of social media sites has exploded, with many focusing on specific interests or certain types of content. Wikipedia lists over 200 active social networking sites.<sup>45</sup> Some have achieved considerable success in little time: within three years of its entry in 2011, Pinterest became the twelfth most visited website in the United States.
- **Search:** Established suppliers include Google, Baidu (the fifth most popular website in the world), Microsoft, Amazon, Yelp, and Travelocity. Examples of recent entry include Duck-

<sup>42</sup> *Statement of Federal Trade Commission* 12, Google/DoubleClick, FTC File No. 071-0170 (Dec. 20, 2007).

<sup>43</sup> *Case COMP/M.7217—Facebook/WhatsApp*, Comm'n Decision, ¶ 188 (Mar. 10, 2014).

<sup>44</sup> *Id.* ¶ 189.

<sup>45</sup> This does not include the hundreds of other social collaborative projects (e.g., Wikipedia), video and photo sharing sites (e.g., YouTube), and virtual social and game worlds (e.g., World of Warcraft, Second Life).



There are many factors

in an online firm's

success, including

investment in

engineering resources,

an attractive user

interface, speed, ease

of use, quality of

content, marketing,

distribution

arrangements, and

complementary

services—not to

mention having a

good idea.

DuckGo (2008), Blekko (2010), Yandex (2010; now the largest search engine in Russia), Apple's Siri (2011), and Facebook's Graph Search (2013).

- **Online advertising:** The Network Advertising Institute, a self-regulatory association for third-party online and mobile advertising, has 96 members that serve interest-based ads or collect similar data for marketing purposes.
- **Data Brokers:** The number of data brokers is unknown but believed to be in the thousands, with new entrants continually offering new types of data and data-derived products.<sup>46</sup>
- **Retail:** Alexa lists over 50,000 online retailers, and six of the top ten lack a brick-and-mortar operation.<sup>47</sup>

A related claim made by some writers is that the volume of data is the key factor determining an online firm's success.<sup>48</sup> This, too, is incorrect. There are many factors in an online firm's success, including investment in engineering resources, an attractive user interface, speed, ease of use, quality of content, marketing, distribution arrangements, and complementary services—not to mention having a good idea.<sup>49</sup> Even extensive stores of data will provide little benefit for a firm unless it has a means of monetizing and innovating based on the data, i.e., talent. In addition, as DuckDuckGo and Apple illustrate, there are many sources of relevant data that can be used as inputs for online services. In short, online firms—whether big or small—can compete through investments in service quality, marketing, or distribution without employing user data.

### Proposed Remedies to Address Alleged Dominance in Big Data Will Harm Competition and Consumers

According to some proponents, competition agencies should impose a variety of restrictions on dominant online firms to reduce the supposedly large advantages they have in terms of access to big data.<sup>50</sup> The proposed restrictions include requiring opt-in consent for collection of data from

<sup>46</sup> *60 Minutes: The Data Brokers: Selling Your Personal Information* (CBS News television broadcast Mar. 9, 2014).

<sup>47</sup> *Shopping Category*, ALEXA (last visited Dec. 7, 2014).

<sup>48</sup> See, e.g., Newman, *supra* note 1, at 403, 407 (referring to big data as “the ‘new oil’ of the information economy” and stating that “entrenched knowledge of consumers’ personal information makes it nearly impossible for any rival or potential rival to woo online advertisers away”); Letter from Mathias Döpfner, Chairman and CEO, Axel Springer SE, to Eric Schmidt, Exec. Chairman, Google Inc. (Apr. 17, 2014) (“If fossil fuels were the fuels of the 20th century, then those of the 21st century are surely data and user profiles.”); Pasquale, *supra* note 36, at 7 (“As long as Google’s search data is secret, no would-be rival will have access to this critical ‘raw material’ for search innovation.”).

<sup>49</sup> Andres V. Lerner, *The Role of “Big Data” in Online Platform Competition* 27–35 (Aug. 26, 2014) (working paper) (“[T]he firm with the most data does not necessarily win, and often does not win.”). See also Lesley Chiou & Catherine Tucker, *Search Engines and Data Retention: Implications for Privacy and Antitrust* (MIT Sloan School Working Paper No. 5094-14) (finding no evidence that larger quantities of historical data confer a competitive advantage for Internet search firms).

<sup>50</sup> See, e.g., European Parliament resolution, *supra* note 1 (calling on the European Commission “to consider proposals aimed at unbundling search engines from other commercial services”); EDPS REPORT, *supra* note 1, ¶¶ 66–68, 72, 80 (outlining proposed remedies); Newman, *supra* note 1, at 446–50 (same); Samuel Miller, *If Google Is a ‘Bad’ Monopoly, What Should Be Done?*, LAW360 (Oct. 22, 2013) (subscription required) (“One remedy that could address this problem is to require Google to enable rivals to access the data that Google has collected about users’ digital interactions for a specified time period (i.e., five years).”); Cédric Argenton & Jens Prüfer, *Search Engine Competition with Network Externalities* 13 (Apr. 13, 2011) (TILEC Discussion Paper No. 2011-024) (“All search engines should be required to share their (anonymized) data on clicking behavior of users following previous search queries among each other and among new entrants[.]”); Martin Cave & Howard Williams, Initiative for a Competitive Online Marketplace, *The Perils of Dominance: Exploring the Economics of Search in the Information Society* (2011) (“[P]ublic policy interventions [could] . . . limit the period over which search data is kept and so narrow the difference in information asymmetries between firms.”).

consumers, providing greater transparency as to the value of collected user data, prohibiting or limiting the collection of user data, divestiture of product lines, and forced sharing of data. Although premised on alleged antitrust violations,<sup>51</sup> most of the proposed remedies appear to be based on a desire to regulate big data or to address perceived shortcomings in the privacy laws.

The antitrust laws were not intended to offer a solution to regulatory- and privacy-based agendas.<sup>52</sup> Instead, the antitrust laws were designed to prevent anticompetitive conduct and to restore competitive conditions when a violation occurs. As the Supreme Court explained, “No court should . . . assume the day-to-day controls characteristic of a regulatory agency.”<sup>53</sup>

In addition, the proposed interventions are in tension with the usual requirements for antitrust remedies. All of the proposed remedies would degrade the user experience or service quality and, as a result, would harm consumers—precisely the opposite of what an antitrust remedy is intended to achieve. For example, requiring affirmative user consent before a website may collect data would detract from the user experience by requiring additional user prompts. (This was a principal criticism of Microsoft’s Vista operating system.) Prohibiting or restricting the collection of user data likewise would make it more difficult for firms to improve their services. Divestiture of other product lines would prevent firms from offering personalized services across product lines.<sup>54</sup> The proposed remedies also would disregard the choices of millions of consumers who have opted into the collection of their personal information or who want to obtain multiple services, e.g., social media and email, from a single online company.

Some of the proposed remedies also present significant administrative challenges. Requiring greater transparency as to the value of a user’s data to the service provider would necessitate determining what information must be shared, when it should be shared, how prominently it must be disclosed, and to whom. This information may confuse or mislead consumers and often will not be known to the provider in advance. For example, personal information often is used to sell advertising in near-instantaneous online auctions. Firms do not necessarily know the value of delivering an ad to a particular consumer ahead of time, and the value of personal information can vary widely by user and over time.

There are also difficult administrative issues with the suggestion that dominant online firms should share their user data. For example, with whom should data be shared, how long should this obligation last, how far back must the company go in providing data, under what conditions (e.g.,

*The antitrust laws were not intended to offer a solution to regulatory- and privacy-based agendas.*

<sup>51</sup> These restrictions allegedly are needed to address antitrust violations premised on refusals to provide user data to rivals, collecting additional user data by expanding into new product areas through organic growth, or collecting additional user data through corporate acquisitions.

<sup>52</sup> *United States v. Phila. Nat’l Bank*, 374 U.S. 321, 371 (1963) (stating that because Congress determined that the effect upon competition is the sole criteria to determine whether a merger violates Section 7, an evaluation of other social or economic factors is improper); **Statement of Federal Trade Commission 2**, Google/DoubleClick, FTC File No. 071-0170 (Dec. 20, 2007) (“Not only does the Commission lack legal authority to require conditions to this merger that do not relate to antitrust, regulating the privacy requirements of just one company could itself pose a serious detriment to competition in this vast and rapidly evolving industry.”); **Case COMP/M.7217—Facebook/WhatsApp**, Comm’n Decision, ¶ 164 (Mar. 10, 2014) (“Any privacy-related concerns flowing from the increased concentration of data within the control of Facebook as a result of the Transaction do not fall within the scope of the EU competition law rules but within the scope of the EU data protection rules.”).

<sup>53</sup> *Verizon Commc’ns Inc. v. Law Offices of Curtis V. Trinko, LLP*, 540 U.S. 398, 415 (2004) (quoting Phillip Areeda, *Essential Facilities: An Epithet in Need of Limiting Principles*, 58 ANTITRUST L.J. 841, 853 (1989)).

<sup>54</sup> The Supreme Court has held that divestiture, the “most drastic” antitrust remedy, is best suited for remedying illegal mergers or acquisitions, not for unilateral conduct. See *United States v. Ford Motor Co.*, 405 U.S. 562, 573 (1972). Divestiture is particularly inappropriate for firms that came by their market position legitimately. See *United States v. Microsoft Corp.*, 253 F.3d 34, 106–07 (D.C. Cir. 2001) (en banc).

price, format) should the sharing take place, what restrictions on the use of the data will be placed on the recipients, and who will enforce these obligations?

Another problem with the proposed restrictions is that they duplicate existing options available to consumers. To prevent or limit collection of their personal information, consumers can adjust their privacy settings, use a browser in “private” mode, or sign out of many online services. Consumers also have numerous ways to opt out of targeted advertising. For consumers who want more data sharing, a number of firms already allow users to export their data to rivals.<sup>55</sup> Consumers who want to be better informed about how their information is being collected and used can consult online privacy policies and guidance from privacy and consumer organizations. Consumers who wish to better understand the value of their data to online providers can access this information from numerous publicly available news articles, as well as financial reports from public companies involved in online advertising. In short, consumers already have options.

Some of the proposed restrictions present additional problems. Requiring greater transparency as to an incumbent’s income or costs related to user data may *reduce* competition. As the FTC noted in response to a proposal to require pharmacy benefit managers to provide certain financial information to its customers, “[M]andated disclosures may actually increase prices. . . . Whenever competitors know the actual prices charged by other firms, tacit collusion—and thus higher prices—may be more likely.”<sup>56</sup>

Another problem with requiring firms to share their data is that this will lessen the incentives of rivals to develop their own sources of data. As the Supreme Court observed in *Trinko*, compelling firms “to share the source of their advantage is in some tension with the underlying purpose of antitrust law, since it may lessen the incentive for the monopolist, the rival, or both to invest in those economically beneficial facilities.”<sup>57</sup> This remedy also raises serious privacy concerns. Many consumers would be upset to learn that a regulator was forcing a company to share their personal information with companies with whom the consumers have no relationship. Ironically, forced sharing of user data could violate company privacy policies or consent decrees that several major online firms have entered into with the FTC.

## Conclusion

Some seem to believe that big data is the next big thing in antitrust law. Proponents of this view are destined for disappointment. To those with only a passing knowledge of antitrust law, the possession of big data by large online firms may sound sinister, but this is in large part because proponents of this view mischaracterize its use and because policy makers and the general public do not yet have a firm understanding of big data. In reality, big data is one of many information inputs into the services that online businesses provide. Since the most important companies in the online world tend to self-generate this information, use it internally to improve the very services that generate it in the first place, and do not disseminate it to others, big data usually is neither a “product” in the antitrust sense nor the type of input that businesses need to obtain from others in order to compete.

---

<sup>55</sup> For example, Google’s Data Liberation Front has developed means for users to export their data using Google Takeout. LinkedIn lets users export all their activity and posts, including profiles, with its “Data Archive” feature.

<sup>56</sup> Press Release, Fed. Trade Comm’n, [FTC Staff: California Bill May Raise Prices for Pharmaceuticals](#) (Sept. 10, 2004). See also ABA SECTION OF ANTITRUST LAW, *PREMERGER COORDINATION: THE EMERGING LAW OF GUN JUMPING AND INFORMATION EXCHANGE* (2006) (describing how antitrust law restricts the sharing of competitively valuable information).

<sup>57</sup> *Trinko*, 540 U.S. at 407–08.

Big data only rarely has anything to do with market definition or competitive effects and regardless of how much critics attempt to drum up excitement about it, this fact is unlikely to change. Big data usually is not a market because the most important online businesses are not data brokers, do not sell big data, and therefore cannot create a market capable of antitrust analysis. Big data will rarely create anticompetitive effects for the same reasons. The notion that a company could monopolize or even have “market power” with respect to user data is implausible, given the ubiquitous and non-rivalrous nature of such information. Any competition affected by the use of big data occurs almost entirely according to how well companies *analyze* information, not according to whether and how much they *obtain* it. Thus the only competitive advantage the typical business can obtain through the use of big data is an advantage based on business acumen, which is not an antitrust concern. To the contrary, business acumen is the core of competition itself.

The recent pace of entry in the online world shows that big data creates no durable barriers to entry or any other significant competitive threat. There is no problem in need of a remedy, and just as importantly, the purported remedies suggested by critics would harm or at least alarm consumers: Some would reduce the amount of data that can be used to improve services, while others would expand the sharing of data beyond consumers’ consent. For these reasons, big data as a topic for competition experts is likely to fade in significance, even as—and perhaps, due to the fact that—the amount of data continues to explode and to become a tool used by ever more firms. ●