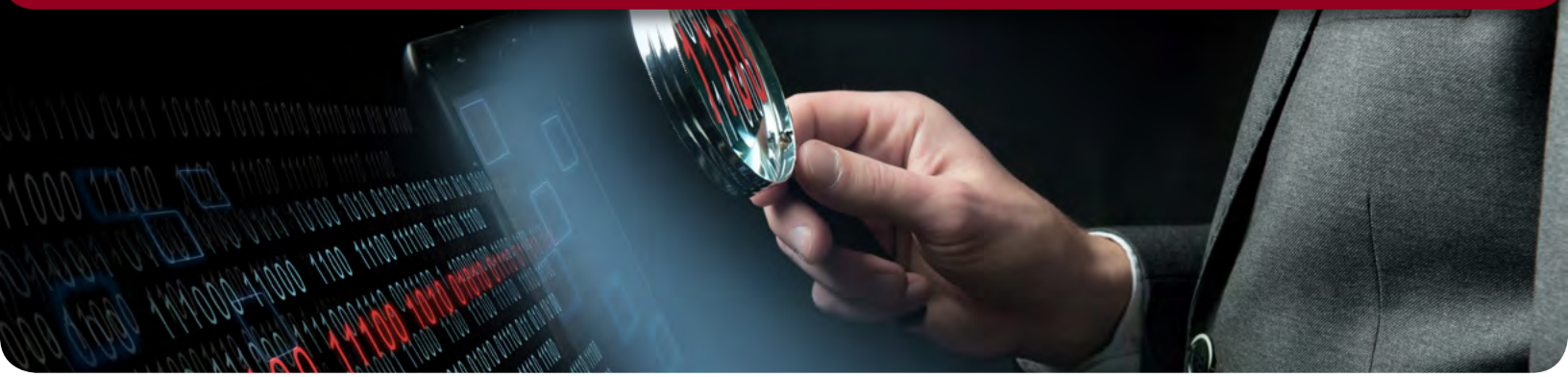
DEFINING "BIG DATA" IN ANTITRUST



BY XAVIER BOUTIN [1] & GEORG CLEMENS [2]

"… smartphones became popular, and Google began offering navigation for free on [Android-powered] phones."

"your competitors come from different fields – they're not the ones you expect them to be. Hilton Hotels would never have expected their margins would be eaten away by Airbnb. Every company starts with a blank sheet of paper, and some of them then arrive like a tsunami."

Corinne Vigreux (TomTom CEO)

I. INTRODUCTION

The story is well known. When it published its decision not to oppose the acquisition of Tele Atlas by TomTom on May 14, 2008, the Commission considered entry by firms offering internet based map applications as "unlikely."[3] Yet, Google soon started to offer navigation for free and TomTom's value slumped from €634m in the last quarter of 2007 to €213m in the first quarter 2009.[4] It is therefore fair to say that the European Commission (the "Commission") and the market in general, underestimated that the volume of data collected by Google and the quality of the collected data ultimately facilitated market entry.

Google's entry in the market for navigable digital map databases in 2009 is one of the earliest and most famous instances of the use of big data for a lateral and drastic entry that would soon revolutionize the entire market. Google Maps Navigation integrated a series of features such as "search along route," "search by voice," traffic view, satellite view and "search in plain English" using the data that was collected from Google's search engine.[5] Moreover, it had the significant advantage

1 Vice President; Compass Lexecon and Université Libre de Bruxelles (ECARES). Email: xboutin@compasslexecon.com. The views expressed in this document do not reflect the opinion of Compass Lexecon, its experts or its clients.

2 Economist; Compass Lexecon. Email: gclemens@compasslexecon.com. The views expressed in this document do not reflect the opinion of Compass Lexecon, its experts or its clients.

3 See Case Comp/M. 4854—*TomTom/Tele Atlas*, Commission Decision of May 14, 2008, pp. 28-29, para 12.

4 See https://www.theguardian.com/business/2015/jul/21/navigating-decline-what-happened-to-tomtom-satnav.

5 See http://googlemobile.blogspot.be/2009/10/announcing-google-maps-navigation-for.html.

of relying on the updates sent by its users. The combination of a high volume and variety of data updated at a high velocity caught incumbents such as TomTom off- guard. In the year following Google's market entry TomTom's revenues plummeted while Google established itself as one of the most important players in the market for navigable digital map databases.

It was then difficult for the Commission, and the market as a whole, to see that cellphones could be used as portable navigation devices (downstream entry), that Google could easily integrate information from the large volume of data it collected such as store locations, restaurant locations and private addresses (downstream data) and, last, that it could easily integrate live traffic data and update maps more quickly than TomTom (upstream entry). This is the reason why Google was not perceived as a potential market entrant by competition experts.

Big data has been a major challenge for most competition authorities and a series of decisions related to big data such as the *Facebook/WhatsApp* merger have faced severe criticism by competition policy experts.[6] The difficulty to tackle antitrust problems in data-related markets is related to the fact that experts have not yet provided a unanimous definition of big data.

This paper fills a gap as it provides a comprehensive definition of big data based on aspects such as volume, variety and velocity commonly associated with big data in IT.[7] There is in our view no such thing as a big data market, or even big data markets. We present big data as a technology. It is therefore a supply related phenomenon that could disrupt practically any market. We will show that the definition of big data provided in this paper allows reconciling diverging opinions in the big data antitrust literature. In addition to that, our definition of big data adequately assesses developments in data-driven markets such as the market for online research. Finally, our big data definition solves the "chicken and egg" problem formulated by Stucke and Grunes (2015).[8]

II. BIG DATA IS NOT JUST A "LARGE VOLUME OF DATA"

While big data has emerged as a much debated competition policy topic, the antitrust literature has failed to provide a unanimous definition of the term big data itself.[9] Commission officials have limited the scope of big data to data volume defining it as "large sets of data."[10] In a similar line, McKinsey defines big data as "large pools of data that can be captured, communicated, aggregated, stored and analysed."[11] This narrow definition of big data has triggered uncertainty among competition policy experts regarding the actual relevance of big data in antitrust. The debate between Tucker and Wellford (2014)[12] and Stucke and Grunes (2016)[13] reflects the ambiguity related to the antitrust relevance of big data.

Tucker and Wellford (2014) argue that big data should not be tackled by antitrust laws.[14] They postulate that big data is characterized by ubiquity, low cost, wide availability and fleeting value. According to Tucker and Wellford these characteristics make big data different from the industry structures which are typically prone to competition problems. Correspondingly Tucker and Wellford neither see big data as a barrier to entry nor as a subject that would raise any other anticompetitive concern.

6 See Stucke & Grunes, "Big Data and Competition," Oxford University Press, 2016.

7 See http://www.gartner.com/it-glossary/big-data.

8 Stucke & Grunes (2015) outline that some companies need enough volume of user data to increase the quality of their data-driven product. Yet large user volumes are only generated if the quality of the product attracts enough consumers. This generates a "chicken and egg" dilemma (See Stucke & Grunes, "Debunking the Myths Over Big Data and Antitrust," CPI Antitrust Chronicle, May 2015).

9 Stucke & Grunes (2016) outline that big data has "many definitions" which are "broad and inclusive" (See Stucke & Grunes "Big Data and Competition," Oxford University Press, 2016). The Autorité de la Concurrence and the Bundeskartellamt refrain providing a definition of big data in their joint paper on competition law in data-driven businesses (See Autorité de la Concurrence and Bundeskartellamt "Competition Law and Data," 2016).

10 Ocello, Sjödin, & Subocs, "What's Up with Merger Control in the Digital Sector? Lessons from the Facebook/WhatsApp EU Merger Case," 1 Competition Merger Brief, 2015, page 6.

11 McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity," 2011.

12 Tucker & Wellford, "Big Mistakes Regarding Big data," the antitrust source, 2014.

13 Stucke & Grunes "Big Data and Competition," Oxford University Press, 2016.

14 Tucker & Wellford, "Big Mistakes Regarding Big data," the antitrust source, 2014.

The claim formulated by Tucker and Wellford on the ubiquity, the low cost, the wide availability and the fleeting value of big data is clearly based on the author's understanding that big data equates to a large volume of data. In their assessment of big data as a barrier to entry the authors analyze whether new entrants are at a severe disadvantage with regards to data accessibility.[15] Hence, Tucker and Wellford exclusively focus on access to data volume. This approach contrasts to the comprehensive definition of big data given by the authors, where big data is not only limited to the collection of data volume but also includes the analysis of the collected data.[16] Paradoxically, Tucker and Wellford do not consider the role of "data analysis" in their assessment thereby possibly underestimating the antitrust relevance of big data.

As opposed to Tucker and Wellford Stucke and Grunes (2015, 2016) classify big data as a potential source for antitrust concerns.[17] Since Stucke and Grunes provide a response to Tucker and Wellford, their arguments necessarily revolve around the same volume-based definition of big data outlined in Tucker and Wellford. Although the authors acknowledge that big data is not only characterized by volume but instead by the four "Vs,"[18] they do not provide a clear-cut definition either. Our approach takes into consideration the most significant characteristics which are volume, velocity and variety, thereby accounting for the dynamics in data-driven markets.

III. DEFINING BIG DATA IN ANTITRUST

Our definition of big data is rooted in the analyses of data-driven markets and cases in data related industries. The OECD recognizes that a big data definition with emphasis on volume alone can be misleading, no matter if the volume is measured in gigabytes, petabytes (millions of gigabytes), or Exabyte (billions of gigabytes).[19] Although the OECD ultimately decided not to provide an alternative definition of big data, it outlined the importance of including concepts such as "velocity" and "variety" which cover more than the mere quantitative aspect of data. Outside of the antitrust debate big data has been commonly characterized by the following three "Vs" which represent the dynamic nature of big data:[20]

- Volume: the volume of data plays a crucial role for data orientated businesses;

- Velocity: …refers to the velocity of data generation but also to the access to data and the way data is processed and analyzed. Velocity adds a dynamic component to the definition of big data as it clarifies that data can quickly lose its value if it is not processed and analyzed quickly. As data loses its value over time, there is a necessity to update and constantly collect data generated by consumers, whether this is the address, the purchase behavior or the internet content; and

- Variety: …refers to the diversity of information data may contain. Data oriented businesses may not only be interested in collecting one type of data from a series of consumers. They potentially aim to elicit a series of data from one type of customers. Yet variety also reflects also the variety of sources data may come from, so that companies may collect data using different platforms.

Velocity requires a data-driven business to update the data regularly and to be able to analyze the data set quickly. The velocity of analyzing data is not only determined by technological prerequisites. Data-driven businesses also have to match the competitor's velocity of data collection and their capability to analyze the data. A company that either manages to process

15 Id. page 7.

16 Id. page 3.

17 Stucke & Grunes, "No Mistake About It: the Important Role of Antitrust in the Era of Big Data," the antitrust source, 2015 and Stucke & Grunes "Big Data and Competition," Oxford University Press, 2016.

18 As we will clarify later, big data can be characterized by three "Vs". These "Vs" are "Volume," "Variety" and "Velocity." Note that some definitions include two more "Vs" which are "Veracity" and "Value."

19 See OECD, "Data-driven Innovation for Growth and Well-being," 2014.

20 These three Vs of big data are sometimes extended to up to four or five "Vs" including characteristics such as "Value" or "Veracity."

and analyze data more quickly than its competitors or to better use the variety of information data can provide, may provide itself a significant competitive advantage. This can be shown using the example of the market for navigable map databases.

In the 2008 *TomTom/Tele Atlas* decision, the Commission excluded the possibility that Google could enter the market for navigable map databases quickly. It argued that the production of a navigable digital map database using end-user feedback was impossible. The Commission claimed that only a certain type of data could be used for navigable digital map databases and that this data needed to be collected through field surveys using customized vehicles.[21] Moreover, it was pointed out that the vast resources needed in building up a navigable digital map database would make entry very costly. Correspondingly, the Commission considered that entry would be unlikely to occur in the next three years.[22]

In 2009, one year after the *TomTom/Tele Atlas* decision, Google announced Google maps navigation for Android and outlined that it was using a navigable digital map database using end-user feedback.[23] Google's ability to compete with internet-based navigable digital map databases led to an upheaval of the market. TomTom which initially failed to react to Google's entry[24] suffered heavily from the competitive pressure that an internet-based navigation provider exerted and it ultimately had to reconsider its entire business strategy.[25]

The Commission clearly based its market entry assessment on a very "volume centered" definition of access to navigable digital map data. According to the Commission, market entry crucially hinged on the ability to duplicate exactly the same data that Tele Atlas and NAVTEQ generated. Moreover, the Commission assumed that exactly the same technology had to be used for this purpose.[26] The Commission did neither consider the fact that Google was able to use the variety of data which its search engine provided, nor did the Commission take into consideration that the search engine produced these data at a very high velocity. In contrast, the volume of data TomTom obtained through the acquisition of Tele Atlas lost its value quickly as €1bn on the Tele Atlas purchase had to be written off in 2009, only one year after the merger.[27]

A series of conclusions can be drawn from the development of the market for digital map databases: First, Tele Atlas and NAVTEQ were not only the dominant market players because of the volume of data they already possessed. Until Google's market entry, no competitor had a technology that could update, analyze and enlarge a high volume of data like Tele Atlas's and NAVTEQ's technology did. Second, Google managed to enter the market because the internet-based navigable maps it used included a variety of data that could be incorporated in the map product. Furthermore it allowed for quicker updates and improvements providing a product that was unmatched by TomTom. Third, in spite of having a head start with respect to the data volume for navigable maps, TomTom could not prevent the successful market entry of Google in the market for navigable map datasets.

In light of these conclusions, we advocate for a departure from the volume orientated approach to big data. Our definition of big data reads as follows: The ability to collect and analyze a large volume of data which contains a variety of information in a timely manner. This definition not only takes into consideration the three "Vs" of big data but also adds a dynamic component to markets that are unmistakably dynamic in their nature. Furthermore, it allows reconciling the critique formulated by Tucker and Wellford and the respective response by Stucke and Grunes and to solve the "chicken and egg" problem for big data.

21 See Case Comp/M. 4854—*TomTom/Tele Atlas,* Commission Decision of May 14, 2008, pp. 28-29, para 12.

22 Ibid.

23 See footnote 3. Note that Google maps navigation entered the market in 2010 http://www.frandroid.com/android/applications/20696_google-maps-navigation-disponible-en-france (Retrieved: March, 21st 2017).

24 See https://www.ft.com/content/01e02434-1cb4-11df-8d8e-00144feab49a.

25 See: https://www.ft.com/content/9e6a2d6a-a163-11e0-baa8-00144feabdc0 and footnote 2.

26 See Case Comp/M. 4854—*TomTom/Tele Atlas,* Commission Decision of May 14, 2008, pp. 28-29, para 12.

27 See footnote 4.

IV. RECONCILING OPPOSING VIEWS ON BIG DATA IN ANTITRUST

As outlined above, Tucker and Wellford focus on the access to consumer data in their assessment and in particular to what extent limited data accessibility may become a barrier to entry. Moreover, the definition of a market for big data is questioned as, according to the authors, a market should only be recognized as such when a product is sold. Tucker and Wellford identify a series of big data characteristics which are its ubiquity, the low cost, the wide availability and the fleeting value of data. Moreover, Tucker and Wellford claim that these characteristics make big data not as prone to antitrust infringements, as other industries.

The characteristics outlined by Tucker and Wellford are not observed for our definition of big data in the antitrust context. The ability to collect and to analyze a large volume of data which contains a variety of information in a timely manner is not ubiquitous. Microsoft and Yahoo unsuccessfully tried to bundle their resources to match Google's search engine abilities but failed to catch up with regards to its performance.[28] Myspace was cautious regarding the collection of private data and therefore failed to quickly implement a successful big data strategy.[29] In contrast to MySpace, Facebook swiftly implemented a big data strategy where a variety of personal data was regularly collected, analyzed and enlarged. This helped Facebook to figure out what the consumers wanted and to swiftly implement new services.[30] Myspace never managed to compete on an equal footing with Facebook after 2008, as it failed to develop a competitive technology to analyze and enlarge a high variety and volume of data in a timely manner, in spite of its initial head start with regards to the possession of large volumes of data.

The technology, which allows collecting and analyzing a large volume and variety of data in a timely manner, is neither accessible at low costs, nor is it widely available. The Commission found in its *Microsoft/Yahoo* decision that large R&D and investment costs are associated with internet search services.[31] Yahoo clarified that heavy investments in software and in hardware are mandatory to stay competitive in the search engine market and that in spite of being significant, Yahoo's investments have only been a fraction of Google's investments.[32] Rough estimates suggest that Google had to invest approximately $585m to roll out its social network Google+.[33]

We agree with Tucker and Wellford that the mere procession of a large volume of data rarely justifies stronger antitrust scrutiny. Yet, companies such as Google, Facebook and Amazon emerge as strong players not because of the volume of data they collected so far. It is their technological ability to harvest data which ultimately provided these companies the dominant position they hold. We therefore also agree with Stucke and Grunes (2016) that big data, as effective data processing technologies, is a relevant antitrust subject. Yet we argue for a more comprehensive definition of big data. The "chicken and egg" problem formulated by Stucke and Grunes vanishes if our definition of big data is adopted.

28 See Comp/M. 5727—*Microsoft/Yahoo! Search Business*, Decision of February 18, 2010.

29 See: http://www.businessinsider.com/former-myspace-ceo-explains-why-facebook-was-able-to-dominate-social-media-despite-coming-second-2015-5?IR=T.

30 "If you have an idea for networking on something, Facebook pushed its tech folks to make it happen. And they kept listening. And looking within the comments for what would be the next application – the next promotion – the next revision that would lead to more uses, more users and more growth"(See: http://www.forbes.com/sites/adamhartung/2011/01/14/why-facebook-beat-myspace/#70195d7e7023, and http://fortune.com/2010/11/19/how-facebook-learned-from-myspaces-mistakes/.

31 See Comp/M. 5727—*Microsoft/Yahoo! Search Business*, Decision of February 18, 2010, para 33.

32 Id. para 144.

33 See: http://www.forbes.com/sites/bruceupbin/2011/06/30/google-cost-585-million-to-build-or-what-rupert-paid-for-myspace/#b9e3b16349f8.

V. THE "CHICKEN AND EGG" DILEMMA IN BIG DATA

In their description of the *Google/Waze* merger, Stucke and Grunes (2015) discern a potential chicken and egg dilemma.[34] In *Google/Waze* the OFT found that Waze's navigation service quality crucially depended on the amount of information it collected from its customers.[35] The more customers use Waze, the better Waze becomes because of the amount of information it collects from its users. Yet, customers would only start using Waze if the quality is good enough.[36] This assessment yields a chicken and egg dilemma that hinges on the fact that large data volumes are defined as an essential input for competition in data-driven markets. Ultimately the OFT concluded that customers would not switch to Waze's service unless the quality of the service was high enough yet can only be safeguarded if enough customers provide their feedback to Waze.[37] Correspondingly, Waze was not seen as a competitive threat for Google by the OFT.[38]

The chicken and egg dilemma tends to overstate the role of data volume for data driven industries. While feedback was crucial for Waze to increase the quality of its navigation service, the same cannot be said for other data-driven industries. Google managed to maneuver out of the dilemma Waze was trapped in because it had a data collection and analysis technology which was superior to Yahoo's technology. Yahoo could not profit from having a higher volume of data than Google and ultimately lost its first spot in the market for search engines. Facebook managed to outperform Myspace, even though Myspace initially possessed significantly more customer data than Facebook.

The above mentioned examples show that volume of data alone may neither prevent market entry nor facilitate market foreclosure. The example of Google's successful entry in the market for navigational map services in 2009 proves that there is not a chicken and egg dilemma *per se* in data-related markets. Google managed to limit Waze's competitive impact because it could rely on a superior technology for data collection and data analysis. Waze could only have become a successful competitor to Google if it had access to an equally performant technology for the collection, analysis and enlargement of data.[39]

The chicken and egg dilemma is based on the assumption that the possession of large data volumes and the processing of data are unrelated. This same idea is present when referring to data as the "oil" of the 21st century, which tends to misrepresent the value of historical data. In our view, data is not comparable to oil; it is rather comparable to wind. Data flows and is largely accessible. Just like wind, it needs to be captured to be transformed into something valuable. Both windmills and data processing technologies certainly improve when tested in real conditions. However, in this context, it is the accumulated experience that is valuable, and not the accumulated wind, nor the historical data.

Therefore, evidence from data-related markets shows that the possession of large data volumes alone rarely causes a problem at all. It is the inability to collect and analyze a large volume of data, which contains a variety of information in a timely manner that yields firms failure in the market.

Moreover, another particularity of these technologies is their incredible versatility. No player in these industries realized that urban transportation or food deliveries could become the data driven industries they are now. There is therefore no closed list of data driven markets, and even less a market for big data itself. Big data is not a demand related phenomenon, but a supply one. Being able to develop the most efficient technology might be more important for success than decades of experience in a traditional sector. We already quoted Corinne Vigreux, TomTom CEO, in the opening of this paper: "Every

34 Stucke & Grunes, "No Mistake About It: the Important Role of Antitrust in the Era of Big Data," the antitrust source, 2015.

35 Office of Fair Trading, Completed acquisition by Motorola Mobility Holding (Google, Inc.) of Waze Mobile Limited, ME/6167/13, para 88.

36 Stucke & Grunes, "No Mistake About It: the Important Role of Antitrust in the Era of Big Data," the antitrust source, 2015.

37 Office of Fair Trading, Completed acquisition by Motorola Mobility Holding (Google, Inc.) of Waze Mobile Limited, ME/6167/13, para 88.

38 Id. para 89.

39 Note however, that Waze might have had the potential to disrupt the market for navigational map services, as it experienced some significant growth, according to the OFT (Office of Fair Trading, Completed acquisition by Motorola Mobility Holding (Google, Inc.) of Waze Mobile Limited, ME/6167/13, para 41 and 42).The OFT therefore potentially disregarded Waze's ability to become a "maverick" in the market.

company starts with a blank sheet of paper, and some of them then arrive like a tsunami." These tsunamis normally don't waste time wondering whether the egg or the chicken comes first.

VI. CONCLUSION

Big data has emerged as a much debated issue in antitrust and is likely to gain additional attention following the steps undertaken by the Commission to foster a European data economy.[40] Yet the antitrust literature has not provided a unanimous definition of big data thereby generating confusion with regards to its relevance. This paper fills a gap as it provides a comprehensive definition of big data based on standard big data concepts from IT. Our definition provides clarification on the relevance of big data in antitrust. Moreover it shows that the chicken and egg dilemma formulated by Stucke and Grunes (2015) becomes irrelevant when our big data definition is applied.

A correct definition of big data is crucial in a correct assessment of data-driven cases by antitrust authorities. In light of the increasing importance of big data strategies for businesses the number of antitrust cases in data related markets is expected to increase in the next years.[41] Antitrust authorities can therefore ill-afford misinterpreting the potential of market entry by companies with a big data strategy as this would lead them to start investigations with the wrong focus. This contribution should therefore be seen as a step to spark a debate on the necessity for a focus in antitrust cases related to big data.

Recent developments in data-driven markets confirm the correctness of our big data definition. Facebook and Google would both be classified as companies with a big data strategy. With the launch of Google+, the closeness of competition between both companies has been unmistakably demonstrated. Google and Facebook are both developing autonomous cars, showing that digitalization may transform a classical industry such as automobile into a big data industry. The increasing importance of big data has been recognized by the Commission, which has taken the next steps towards a European data economy.[42] Our contribution therefore provides an important tool for the identification of market entry and potential competitors in future antitrust cases related to data-driven markets.

40 See European Commission, "Commission outlines next steps towards a European data Economy," Brussels, January 10, 2017.

41 Only 1.8 percent of the firms that participated in the New Vantage Partner LLC "Big Data Executive survey" claimed that big data was not important to that firm (See New Vantage Partners "Big Data Executive Survey 2016," available at: http://newvantage.com/wp-content/uploads/2016/01/Big-Data-Executive-Survey-2016-Findings-FINAL.pdf.).

42 See footnote 40.