Chapter 1. Big data as a fundament of data-driven online platforms' business models

1. Introduction

The main goal of this chapter is to present big data as a foundation upon which business models of data-driven online platforms are built and clarify basic concepts that structure further analysis. For this purpose, it is necessary to define big data and explain its main characteristics, the so-called "4Vs of big data," i.e. its velocity, variety, volume, and value, which affect online platforms' businesses. Although much has already been said about big data in general and big data in antitrust in particular, definitions proposed in the literature differ significantly, and there is still no agreement on how to define this concept.[1] Since differences in the understanding of big data may affect the assessment of data-driven practices employed by online platforms, it is particularly important to put forward a definition that is used in the book and may determine its conclusions.

To capture the effects that big data may have on competition between online platforms, it is also crucial to present economic characteristics of data itself and distinguish between personal and non-personal data. Determining whether data is ubiquitous, non-rivalrous, non-exclusive, low-cost, widely available, non-substitutable etc. is even more important since it may influence the assessment of data as a barrier to entry or expansion and as an essential facility. Once big data is defined it must be considered whether big data may become a source of a competitive advantage for online platforms. To that end, in the first place, big data value creation cycle is presented, i.e. how raw data can be transformed into value, starting from data generation, collection, processing, analysis, and finishing with data-driven decision-making and monetisation.

[1] Xavier Boutin and Georg Clemens, "Defining 'big data' in antitrust," *CPI Antitrust Chronicle* August 2017: 1; OECD, "Data-driven Innovation for Growth," 11.

Big data is then examined as a source of efficiencies and benefits, not only for businesses, but also for consumers and society in general. Next, it is considered whether big data should be perceived as an input of production to products or services offered by online platforms or rather as a commodity that can be traded. Based on economic characteristics of data, it is also analysed whether big data can constitute a barrier to entry and expansion for potential or current competitors, thereby potentially reinforce the competitive advantage enjoyed by an incumbent. Finally, it is examined to what extent big data can constitute a competitive advantage for data-driven online platforms.

2. Concept and characteristics of big data

Big data has undeniably become a buzzword that appears throughout the literature regardless of its subject matter and field of science. Thus, it seems surprising that so far, the concept lacks a standard definition.[2] Therefore, this section provides a comprehensive definition of big data that involves the so-called "4Vs of big data," i.e. its volume, velocity, variety, and value. However, to structure the debate about big data's role in online platforms' businesses and competition among them, it is not only crucial to have a common understanding of what is meant by big data, but also what economic characteristics the underlying data has, what the difference between personal and non-personal data is and what implications that distinction may have on the assessment of unilateral practices of incumbent platforms.

2.1. Definition of big data – the "4Vs of big data"

Definitions proposed in the literature, regardless of which branch of science or field of research, differ significantly.[3] Even between antitrust scholars, there is no agreement on how this concept should be understood, which may lead to misunderstandings and improper assessment of its actual effect on competition between online platforms.

[2] See e.g.: Gil Press, "12 Big Data Definitions: What's Yours?," Forbes, 3 September 2014, available at http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours; MIT Technology Review, "The Big Data Conundrum: How to Define It?", 3 October 2013, available at http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/.

[3] De Mauro, Greco and Grimaldi, "A Formal Definition," 128.

A large fraction of literature focuses mainly on the volume of data,[4] indicating that big data refers to "large sets of data,"[5] "large amount of data"[6] or "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse."[7] However, as rightly indicated by the OECD, focusing on the volume alone can be misleading since volume is only one of the relevant features which shape big data phenomenon,[8] next to its velocity, variety, and value.[9] Definitions based solely on the volume of data miss the actual essence of big data – the importance of the underlying technology, which allows for value to be extracted from raw data. Large datasets without adequate technology to process them would remain just large datasets.[10]

According to Tucker and Wellford, "big data refers to a collection of data sets so large and complex that traditional database systems cannot effectively manage or process the information."[11] Their definition focuses on the volume of big data, as well as the fact that traditional technologies might not be able to process and manage it, but disregards the other key features of big data, i.e. its velocity, variety, and value. The Commission regards big data as one of its policy priorities with the Digital Single Market strategy and defines big data as "large amounts of data produced very quickly by a high number of diverse sources."[12] The definition proposed by the Commission reflects volume, velocity,

[4] See e.g. Daniel L. Rubinfeld and Michal Gal, "Access Barriers to Big Data," *Arizona Law Review* 2017, Vol. 59: 345.

[5] Eleonora Ocello, Cristina Sjödin and Anatoly Subocs, "What's Up with Merger Control in the Digital Sector? Lessons from the Facebook/WhatsApp EU Merger Case," *Competition Merger Brief* 2015, No. 1: 6.

[6] Andres V. Lerner, "The Role of 'Big Data' in Online Platform Competition," 3, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780.

[7] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela Hung Beyers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," 1, available at https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx.

[8] OECD, "Data-driven Innovation for Growth," 11.

[9] OECD, *Supporting Investment in Knowledge Capital, Growth and Innovation* (OECD Publishing: 2013), 324–325.

[10] Hu Han, Yonggang Wen, Tat-Seng Chua and Xuelong Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," *IEEE Access* 2014, Vol. 2, 654, available at http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6842585.

[11] Darren S. Tucker and Hill B. Welford, "Big Mistakes Regarding Big Data," *The Antitrust Source* (December 2014), 2, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2549044.

[12] European Commission, "Digital Single Market strategy: Big data," available at https://ec.europa.eu/digital-single-market/en/policies/big-data; European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions — Towards a thriving data-driven economy," COM/2014/0442 final, 4.

and variety of big data, but disregards its other crucial feature, i.e. its value. An interesting definition has also been proposed by Boutin and Clemens, who consider big data as "the ability to collect and analyse a large volume of data which contains a variety of information in a timely manner."[13] Their definition perceives big data as a specific ability and focuses mainly on a technology that enables processing and analysis of large datasets. Such a definition, however, seems to disregard the question of the access to data and the fact that big data is primarily an asset or an input. Undoubtedly, appropriate technology is crucial for the analysis of data, but even if an undertaking owns adequate infrastructure to process large amounts of data, it first needs to have access to the relevant data, and, as it is shown below, not all data is easily available.

Moreover, some authors, although elaborating on the effects of big data on the competition law, do not provide an unambiguous definition of this term. For example, Stucke and Grunes seem to agree that big data should be characterised by the 4Vs,[14] but they do not define big data and simply indicate that it has "many definitions," which are "broad and inclusive."[15] A similar approach was adopted by the Autorité de la Concurrence and the Bundeskartellamt in their joint report "Competition law and data."[16]

As demonstrated above, there are significant differences as regards the meaning of the big data in competition law literature. This may, in turn, lead to misconceptions regarding the actual impact of big data on the competition between online platforms and the assessment of their potentially anticompetitive behaviour.[17]

Taking into consideration all the above-mentioned definitions and remarks, in the view of the author, a definition that best reflects the actual meaning of big data is the one proposed by De Mauro et al.: "big data is the information asset characterised by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value."[18] This definition simultaneously envisages big data's function as an asset and the

[13] Boutin and Clemens, "Defining 'big data'," 4.

[14] Maurice E. Stucke and Allen P. Grunes, "No Mistake About It: the Important Role of Antitrust in the Era of Big Data," *The Antitrust Source* (April 2015), 2–3, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2600051.

[15] Stucke and Grunes, *Big Data*, 15.

[16] Autorité de la Concurrence and Bundeskartellamt, "Report – Competition Law," 4–5.

[17] It can be demonstrated by the discussion between Stucke and Grunes on the one side, and Tucker and Welford on the other. Their different perception of big data led them to different conclusions with regard to big data as a potential source for antitrust concerns. See: Stucke and Grunes, "No Mistake"; Tucker and Welford, "Big Mistakes."

[18] De Mauro, Greco and Grimaldi, "A Formal Definition," 131.

role of technology in its analysis and transformation into value. It also reflects the distinctive features of big data, which are pointed out in the literature, in particular the OECD and the Commission's documents, i.e. big data's volume, velocity, variety, and value.[19] The first three "Vs" (volume, variety, and velocity) appear in virtually all definitions and are considered key elements of big data's definition.[20] The fourth "V," which refers to the value that can be extracted from big data,[21] is accentuated, especially in business and antitrust literature.[22] Each of these features is discussed in turn.

2.1.1. Volume

The large volume of data is probably the most self-evident feature of big data, which is indicated in majority of the proposed definitions.[23] It refers to the large amount of data that is collected, analysed, and further utilised.[24] Nonetheless, it should be noted that it is not possible to establish what is meant by the "large" volume of data in isolation from circumstances of the case at issue. The question of what constitutes a large volume of data may vary by industry, business model, size of the market or even by region. Moreover, the amount of data that is currently is considered "large" may not be considered as such in the future. Thus, it seems purposeless to define the exact amount of data, which determines whether a specific dataset amounts to big data, or not.[25] In principle, one could argue that large datasets should be understood as datasets, whose

[19] Han, Wen, Chua and Li, "Toward Scalable Systems," 654; Autorité de la Concurrence and Bundeskartellamt, "Report – Competition Law," 4.

[20] See e.g. Douglas Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety, Meta Group (Gartners Blog post)," available at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf; OECD, "Data-driven Innovation for Growth," 11.

[21] OECD, *Supporting Investment*, 322; Stucke and Grunes, *Big Data*.

[22] Other authors either disregard the "value" of big data and add "veracity" as the fourth "V," or include both value and veracity, thereby creating 5 "Vs" of big data. See e.g.: European Commission, Commission Staff Working Document on Online Platforms, 6; Roberto Moro Visconti, Alberto Larocca and Michele Marconi, "Big Data-driven Value Chains and Digital Platforms: From Value Co-creation to Monetization," available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903799. Veracity refers to trustworthiness of data and relies on the assumption that all data generated on the internet provides trustworthy information. However, since trustworthiness of data may raise controversies, it is excluded from the proposed definition.

[23] OECD, "Data-driven Innovation for Growth," 11.

[24] Boutin and Clemens, "Defining 'big data'," 3; Nir Kshetri, "Big data's impact on privacy, security and consumer welfare," *Telecommunications Policy* 2014, Vol. 38(11): 1138.

[25] By the same token, it might be pointless to state that X amount of data is indispensable to operate business Y.

"size is beyond the ability of typical database software tools to capture, store, manage and analyse."[26]

In this context, it is worth mentioning that the volume of data processed globally is increasing at an unprecedented rate,[27] and is forecast to grow from 3.4 zettabytes per year in 2014 to 10.4 zettabytes by the end of 2019.[28] It is estimated that by the year 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet.[29] Such a huge increase in the volume of data can be attributed to the digital shift, in particular the extensive use of multiple smart devices equipped with digital sensors (e.g. smartphones, computers, tablets, GPS locators[30]) and the rise of online platforms. It should also be indicated that almost all media, social, economic, and official activities have moved to the online world (e.g. e-commerce or e-government services).[31] As a result, even traditionally "offline" businesses or institutions are compelled to switch to the internet to keep pace with technological developments and to satisfy the needs of the new generation of online customers, which increases the amount of data in circulation even further.

Although generally the access to data has been facilitated and the costs of data collection, storage and analysis have decreased,[32] the growing volume of data continues to pose challenges to its storage and analysis.[33] Actually, to effectively manage, store and analyse such large datasets undertakings are either forced to develop new technologies by themselves, or outsource storage and

[26] Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Hung Beyers, "Big Data," 1; Edd Dumbill, "Making Sense of Big Data," *Big Data* 2013, Vol. 1(1): 1.

[27] IBM, "Analytics: The real-world use of big data. How innovative enterprises extract value from uncertain data," 4, available at https://public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03519usen/GBE03519USEN.PDF.

[28] Cisco Global Cloud Index, "Forecast and Methodology, 2014–2019," available at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index- gci/Cloud_Index_White_Paper.html.

[29] Bernard Marr, "Big Data: 20 Mind-Boggling Facts Everyone Must Read," 30 September 2015, available at https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#66035c6f17b1.

[30] De Mauro, Greco and Grimaldi, "A Formal Definition," 125.

[31] OECD, "Big data," 6.

[32] OECD, "Big data," 6; Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Hung Beyers, "Big Data," 2. The OECD indicates that these developments constitute realisation of Moore's Law, according to which the number of transistors that can be located on integrated circuits doubles about every two years since their invention, therefore increasing efficiency of electronic devices and decreasing their cost over time. In other words, the amount of computing power that can be obtained for the same amount of money doubles approx. every two years.

[33] The OECD gives an illustrative example for problems related to storage of such amount of data. Namely, in order to store 10.4 zetabytes of data, each and every individual in the world (including infants) would have to own eleven iPhones of 128 gigabytes. See: OECD, "Big data," 6.

processing services from external providers (e.g. cloud service providers in case of data storage). Notwithstanding this, the management of the big data's growing volume requires significant and constant investment.

2.1.2. Velocity

This feature refers to the speed with which data is generated, accessed, processed, as well as used. As the technology of data analysis advances, some undertakings are able to process and utilise data at speeds approaching real time.[34] The velocity of big data is closely related to another feature of big data, which is big data's timeliness or time-sensitivity.[35] This feature implies that some data is time-sensitive and may lose value if it is not stored, processed, and analysed quickly. In dynamic markets, such as those on which online platforms operate, new data might render some old data out-dated.[36] Moreover, certain kinds of time-sensitive data must to be analysed in close to real time to be valuable for a business.[37] Therefore, in some markets and industries, data may need to be regularly updated or collected anew, as the data obtained with a lag may be valueless.[38] For example, information concerning traffic or accidents is valuable for a road-map applications only when it can be accessed in a timely manner; the data about a consumer's current location is valuable for advertisers only if they can display him or her an ad with "the best restaurant in area" exactly when he or she is in that area; if a consumer searches for new headphones now, there is no use to show him or her an advertisement of headphones in a month's time, as he or she could have already purchased it.

[34] IBM, "Analytics: The real-world," 4. This phenomenon known as now-casting describes the ability to observe an event happening now and using it to predict things as they occur (e.g. notifying outbreak of influenza based on observed increase in search queries concerning remedies for flu). Although now-casting concept is associated with meteorology and weather prediction, it becomes popular with other industries, especially online, as it allows to predict real-time market trends in various sectors. See: OECD, "Big data," 6, 8; Stucke and Grunes, *Big Data*, 19.

[35] Kshetri, "Big data's impact," 1138.

[36] Rubinfeld and Gal, "Access Barriers," 347.

[37] IBM, "Analytics: The real-world," 5.

[38] Some data may lose its value over time but it does not mean that all data share this characteristic. In certain cases, historical data may be equally valuable, e.g. analysed in order to identify certain trends. See: Tucker and Welford, "Big Mistakes," 4.

2.1.3. Variety

Variety refers to the diversity of data that is collected (personal and non-personal data)[39] and the variety of information that can be derived from data.[40] It also reflects the wide array of sources which data may be derived from (data may originate from users, machines, computers, different platforms, etc.), the variety of forms that data can take (such as text, pictures, videos, audios, click streams or voice[41]), as well as different formats that data can have (structured, semi-structured and unstructured data[42]). Such diversification of data requires the application of a wide range of tools and creating new ones to effectively manage its complexity.[43] Moreover, the same data can have many different applications and be valuable for multiple businesses or industries.

The distinction between structured, semi-structured or unstructured data is particularly important in that regard.[44] Structured data encompasses information, mostly text files, with a high degree of organisation, usually displayed in titled columns and rows, which can be easily accessed, labelled and processed by organisations using traditional data mining tools[45] (e.g. consumer data listed in alphabetical order, sales ledger or other registers that are well-organised and thus relatively easy to process). In contrast, unstructured data is a huge compilation of various data that has no identifiable internal structure and which is generally valueless until identified, organised in an orderly manner and processed adequately. Unstructured data typically encompasses the human-generated and people-oriented content that does not conform to specific database tables or models.[46] As it is more difficult to analyse than

[39] Autorité de la Concurrence and Bundeskartellamt, "Report – Competition Law," 5.

[40] Analysis of one piece of data may lead to obtaining variety of information, which may in turn have multiple applications.

[41] IBM, "Analytics: The real-world," 4

[42] IBM, "Analytics: The real-world," 4; Autorité de la Concurrence and Bundeskartellamt, "Report – Competition Law," 6; OECD, *Data-driven Innovation: Big Data for Growth and Well-Being* (OECD Publishing: 2015), 151.

[43] IBM, "Analytics: The real-world," 4; Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Hung Beyers, "Big Data," 83.

[44] Autorité de la Concurrence and Bundeskartellamt, "Report – Competition Law," 6; OECD, *Data-driven Innovation*, 151.

[45] BrightPlanet, "Structured vs. Unstructured data," available at https://brightplanet.com/2012/06/structured-vs-unstructured-data/; Sherpasoftware, "Structured and Unstructured Data: What is It?," available at http://sherpasoftware.com/blog/structured-and-unstructured-data-what-is-it/.

[46] Jim Harris, "Bridging the Divide between Unstructured and Structured Data," available at https://datascience.berkeley.edu/structured-unstructured-data/; Darin Stewart, "Big Content: The Unstructured Side of Big Data," available at http://blogs.gartner.com/darin-stewart/2013/05/01/big-content-the-unstructured-side-of-big-data/.

structured data,[47] its processing usually requires an application of more sophisticated and innovative data analytics tools. Examples of unstructured data include emails, pictures, social media posts, videos, audios, voice, spreadsheets or financial transactions.[48] Lastly, data may combine features of these two categories, i.e. it may not conform into a predefined model but have some organisational properties that make it easier to process. Such data is referred to as semi-structured.[49] In this respect, it should be emphasised that currently a majority of commercially valuable data is unstructured, and, therefore, its processing and extracting value from it requires special tools and adequate technology.[50] It is argued that once organised and analysed, unstructured data may provide substantially more valuable insights for an undertaking than structured data.[51]

A good example that illustrates the variety of data and potential data sources are user profiles that are created by online platforms. The huge increase in the number of sensors, smart devices and social collaboration technologies has enabled online platform providers to get access to various data that is voluntarily shared, bought from data brokers, made available by other organizations or collected through other sources to which they have access. Such data may include basic customer data such as an address (physical or IP), date of birth or gender, but also racial or ethnic origin, political opinions, religious or philosophical beliefs, diseases, hobbies, allergies, dietary habits or purchasing history. All such data is used to create a detailed profile of an individual that is further analysed and used for various purposes, e.g. targeted and behavioural advertising, or employing price discrimination mechanisms.[52]

---

[47] Analysis of unstructured data poses many challenges for businesses, such as data quality, data categorisation, merging structured and unstructured data, as well as managing the potentially large amount of information derived from that data. See: Stephen Pritchard, "How to manage unstructured data for business benefit," *Computer Weekly*, 5 October 2012, available at http://www.computerweekly.com/feature/How-to-manage-unstructured-data-for-business-benefit.

[48] Nir Kshetri, "The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns," *Big Data & Society* July–December 2014: 7.

[49] Autorité de la Concurrence and Bundeskartellamt, "Report – Competition Law," 6; OECD, *Data-driven Innovation*, 151.

[50] Autorité de la Concurrence and Bundeskartellamt, " Report – Competition Law," 6; OECD, *Data-driven Innovation*, 151.

[51] OECD, *Data-driven Innovation*, 151; Harris, "Bridging"; Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Houghton Mifflin: Harcourt, 2013), 30–31.

[52] OECD, "Big data." Price discrimination is further discussed in chapter 4.

2.1.4. Value

The above-discussed three basic features reflect the technical properties of big data that rely on technological advancements, especially in data storage and analysis.[53] However, it is arguably the fourth "V" – big data's value – that motivates undertakings to collect, process and use data.[54] Big data's value refers to the social and economic value that can be derived from the utilisation of big data, in other words, the transformation of information included therein into insights that may create economic value for undertakings and benefit the society.[55] Big data's value is considered to be both a cause and a consequence of the increase in volume, variety, and velocity of data.[56]

Notwithstanding this, it should be emphasised that data have no intrinsic value, and its value depends on the context of its use.[57] The value of data usually does not lie in the collected data itself but instead depends on the knowledge that can be extracted from it and its further use.[58] Therefore, specific data might not be equally valuable for all organizations. In fact, data valuable for one undertaking may be useless for others. Most importantly, it should be emphasised that it is the information and knowledge that can be derived from data, and not raw data, that creates value for undertakings.[59] Data in itself is just a raw material that needs to be processed to become valuable.[60] To extract value from data and to fully exploit big data's potential, the assets that are equally or even more necessary than data include an adequate and innovative technology (big data analytics, including machine learning),[61] as well as highly skilled IT staff.[62] In this sense, the mere

[53] OECD, *Supporting Investment*, 325.

[54] OECD, *Supporting Investment*, 325.

[55] De Mauro, Greco and Grimaldi, "A Formal Definition," 131.

[56] OECD, "Big data," 6.

[57] OECD, *Data-driven Innovation*, 197.

[58] Inge Graef, "Market Definition and Market Power in Data: The Case of Online Platforms," *World Competition: Law and Economics Review* 2015, Vol. 38(4): 479; Geoffrey Manne and Ben Sperry, "The Problems and Perils of Bootstrapping Privacy and Data into an Antitrust Framework," *CPI Antitrust Chronicle* May 2015: 9.

[59] Livio Cricelli and Michele Grimaldi, "A dynamic view of knowledge and information: a stock and flow based methodology," *International Journal of Management and Decision Making* 2008, Vol. 9(6): 686–698.

[60] Therefore, Boutin compares data to wind, which flows and is widely available. Data, like wind, has to be captured, processed and transformed into value. In case of wind that would be energy and in case of data – valuable information. See: Boutin and Clemens, "Defining 'big data,'" 6.

[61] See e.g. OECD, *Data-driven Innovation*, 36; De Mauro, Greco and Grimaldi, "A Formal Definition," 125–126; Carl Shapiro and Hal Varian, "Information Rules: A Strategic Guide to the Network Economy," *Harvard Business School Press* 1999: 8; Stucke and Grunes, *Big Data*, 23–24.

[62] Analysing big data at the right speed requires special computational and storage capacities that an average IT system might not be able to grant. See: De Mauro, Greco and Grimaldi, "A Formal Definition," 125.

size of a dataset (i.e. its volume) is not all that matters. Although access to large datasets is indispensable, without proper technology and expertise it remains just a large dataset. The process of transforming raw data into value, i.e. the so-called big data value creation cycle, is discussed further in section 3.1. below.

2.2. General economic characteristics of data

While considering big data from a competition law perspective, it is essential to elaborate on the economic characteristics of data itself. Regardless of whether data is perceived as an input of production or a commodity, there is a consensus that it has specific characteristics that make it different from any other asset owned by an undertaking.

First, by its nature, data is inherently non-rivalrous, which means that consumption of the data does not decrease its availability to the others.[63] In other words, the same data can be collected and used multiple times without losing value and being depleted. In contrast to oil,[64] which is an example of a purely rivalrous good that can be consumed only once, data represents a non-rivalrous good.[65] In principle, data can be consumed and utilized an unlimited number of times and does not use up. Consequently, using a specific dataset by one provider does not prevent other providers from using the same dataset, provided they have access to it.[66] If one undertaking collects data about an individual, such as his or her name, data of birth, gender, postal address, phone number, current interests and hobbies, the fact that it uses such data does not preclude its competitors from accessing and processing the same data. Therefore, data can be used simultaneously as an input factor by different undertakings, repeatedly and for multiple purposes.[67]

However, although data is basically non-rivalrous it does not mean that it is also non-excludable and constitutes a pure public good.[68] In this context, it should be emphasised that data is not a homogenous good. On the one hand, there are categories of data whose possession by one provider does not prevent

[63] OECD, *Data-driven Innovation*, 179.

[64] The Economist, "The world's."

[65] Therefore, comparing data to oil seem rather misguided. See: OECD, "Data-driven Innovation for Growth," 24–25; Charles I. Jones and Christopher Tonetti, "Nonrivalry and the Economics of Data, Working Paper No. 3716," available at http://christophertonetti.com/files/papers/JonesTonetti_DataNonrivalry.pdf.

[66] Graef, "Market Definition," 479.

[67] OECD, *Data-driven Innovation*, 181.

[68] Thus, data cannot be considered a public good, which is both non-excludable and non-rivalrous. See: Schepp and Wambach, "On Big Data," 121.

others from collecting the same piece of data. These categories entail data that is generally available, which users must provide to use most of online services, e.g. profile data (name, surname, age, gender etc.). Thus, this type of data can be assumed non-excludable, as it is not possible to exclude others from accessing and using it.[69] However, on the other hand, there is also data that has been collected by an undertaking as a result of its investment in data collection and analysis (e.g. Google's search algorithm and ranking system). Such data, which is often crucial for operating on a particular market, may not be so easy to obtain by competitors. They would have to incur similar investments as an incumbent, but even then, it is not certain whether that would allow them to acquire similar data and use it in an equally successful manner. If the data is in the exclusive control of an incumbent, the latter will likely refuse to give competitors access to such data. As such, the incumbent may be able to exclude competitors from accessing and using such data,[70] which consequently makes this type of data excludable. Against this background, it is argued that exclusive control over data, particularly user data, exercised by some undertakings, may give rise to insurmountable barriers to entry.[71]

Moreover, data is claimed to be ubiquitous and widely available.[72] That is, however, only partially true. Although data is generated on a massive scale, that may not result in its wide availability. There is data that can be seen as widely available, due to its ubiquity, easiness to collect, relative cheapness[73] and, in case of personal data, willingness to share by users.[74] For example, user's name, email address and age or gender may be ubiquitously shared and easily gathered by various undertakings. It is argued that the reason for wide availability is the fact that individuals regularly leave digital footprints, generate a huge amount of content on the internet, and "multi-home," which means that they use multiple

[69] Robert P. Mahnke, "Big Data as a Barrier to Entry," *CPI Antitrust Chronicle* 2015 (May): 3.

[70] Schepp and Wambach, "On Big Data," 121; Graef, "Market Definition," 479.

[71] See e.g. Newman, "Search, Antitrust and," 401–454.

[72] Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Hung Beyers, "Big Data," 2; The Economist, "Special report on managing information: Data, data everywhere," 25 February 2010, available at https://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf; Tucker and Welford, "Big Mistakes," 3.

[73] It was estimated that basic information about an individual, such as his age, gender and location is worth $0.0005 per person. Even a detailed profile of a person, including his sensitive data or current needs, usually costs less than one dollar. See: Emily Steel, "Financial Worth of Data Comes in at Under a Penny a Piece," Financial Times, 12 July 2013, available at https://www.ft.com/content/3cb056c6-d343-11e2-b3ff-00144feab7de.

[74] Geoffrey Manne and Ben Sperry, "Debunking the Myth of a Data Barrier to Entry for Online Services, Truth on the Market," available at http://truthonthemarket.com/2015/03/26/debunking-the-myth-of-a-data- barrier-to-entry-for-online-services/; Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Hung Beyers, "Big Data," 1–2.

different providers for the same or different services. However, there is also data that cannot be easily and widely obtained or replicated, in particular data that is the result of a specific investment, use of an innovative algorithm or unique data processing techniques.[75] For example, data about user's financial situation, diseases or political beliefs is not so eagerly disclosed and therefore is more difficult to gather. Similarly, it is probably impossible to replicate data that Google has collected about users' search history or data that Amazon has gathered about its users' purchases.[76] Consequently, the latter type of data cannot be perceived as ubiquitous and widely available for all market players. Moreover, it seems that the success of some businesses may be owed to the fact that they have superior access to data that its competitors cannot easily obtain.[77] Notwithstanding this, although in some instances, raw data could be considered widely available, what is scarce is the ability to extract knowledge from it.[78] In this context, it should also be pointed out that data availability is extensively affected by the legal framework governing data collection and use.[79] Therefore, even if from the economic perspective data could be seen as widely available, its processing and use may be significantly restricted by applicable legal provisions, which may affect the actual availability of data.

Furthermore, some argue that data is non-substitutable,[80] which means that only particular data can bring expected results of data and there is no other thing, i.e. a substitute, which can serve the same purpose.[81] It seems that whether data is substitutable or not depends on the circumstances of each case and on a particular business model of an undertaking. For example, while a video-sharing platform (e.g. YouTube) can learn about music preferences of an individual user by observing his or her activity on the platform and the music

[75] Graef, "Market Definition," 483.

[76] The Economist, "Big tech faces competition and privacy concerns in Brussels," 23 March 2019, available at https://www.economist.com/briefing/2019/03/23/big-tech-faces-competition-and-privacy-concerns-in-brussels.

[77] See e.g. Google search and Bing search. It is rather unlikely that Bing search engine would be able to challenge Google's market position without having access to its search data.

[78] The Economist, "Special report."

[79] Marc Bourreau, Alexandre de Streel and Inge Graef, "Big Data and Competition Policy: Market power, personalised pricing and advertising. CERRE report," 8.

[80] In the economic literature substitutability is also referred to as fungibility. See: GSMA, "The Data Value Chain," 9, available at https://www.gsma.com/publicpolicy/wp-content/uploads/2018/07/GSMA_Data_Value_Chain_June_2018.pdf.

[81] R. C. Rockwell, "Funding Social Science Data Archiving and Services in the Networked Environment," *Journal of Library Administration* 1999, Vol. 26(1–2): 89; Anja Lambrecht and Catherine E. Tucker, "Can Big Data Protect a Firm from Competition?," 11–15, available at http://ec.europa.eu/information_society/newsroom/image/document/2016-6/computer_and_communications_industry_association_-_can_big_data_protect_a_firm_from_competition_13846.pdf.

he or she listens to, a social network platform can acquire the same knowledge by analysing his or her profile information and posts that he or she voluntarily shares on its platform.[82] Moreover, although it might be true that under some circumstances data can be non-substitutable, it should be stressed that in case of big data it seems more reasonable to focus on substitutability of the tools used to its analysis, as it is not data itself that contributes to an undertakings' success but a group of other factors, such as innovative data processing techniques and skilled labour force.[83]

Data is also considered low-cost.[84] In that regard, it is commonly argued that data has near-zero marginal cost of production and distribution[85] and technologies used to analyse or host data are also often available for free (e.g. open-source technologies such as Hadoop[86] or Cassandra[87]). Moreover, the costs of collecting big data, its storage and analysis are considered relatively low and gradually decrease.[88] However, this assumption is rarely correct. In most cases, data collection, storage and processing may be costly,[89] and widely available tools do not allow for extracting the expected value from data.[90] This could be illustrated by the fact that according to some estimates the number of data-driven mergers and acquisitions, whereby the incumbent online platforms, such as Google, Facebook, Microsoft or Apple, acquire smaller platforms with rich datasets, more than doubled in recent years[91] and their value reached levels not seen before.[92] Moreover, some data can be purchased, for example, from

[82] Graef, "Market Definition," 479.

[83] There are examples of companies without established data advantage that were nevertheless able to attract more customers than established players and disrupt the industry. Their success is owed not to possessing specific data, but to a superior value proposition. See: Lambrecht and Tucker, "Can Big Data," 11–13.

[84] Executive Office of the President of the United States of America, "Big data: seizing opportunities, preserving values. 2015 Interim Progress Report," 4; Shapiro and Varian, "Information Rules," 3, 24; Daniel D. Sokol and Roisin E. Comerford, "Antitrust and regulating big data," *George Mason Law Review* 2016, Vol. 23(5): 1136.

[85] Shapiro and Varian, "Information Rules," 3, 24.

[86] Hadoop was originally funded by Yahoo, but now is available as an open source solution and drives many of current big data processing platforms. See: OECD, *Data-driven Innovation*, 79.

[87] Tucker and Welford, "Big Mistakes," 3.

[88] Executive Office of the President of the United States, "Big data," 2, 4.

[89] Forbes, "Big cost of big data," 16 April 2012, available at https://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data/#4ab037af5a3b.

[90] Stucke and Grunes, "No Mistake," 7.

[91] European Data Protection Supervisor, "Report of workshop on Privacy, Consumers, Competition and Big Data," available at https://edps.europa.eu/sites/edp/files/publication/14-07-11_edps_report_workshop_big_data_en.pdf.

[92] See e.g. Facebook's acquisition of WhatsApp for 19 billion dollars, or Microsoft acquisition of LinkedIn for 26.2 billion dollars.

data brokers.[93] However, it is also worth mentioning that if data is acquired from data brokers, its cost may differ depending on many variables and it cannot be established upfront that the cost of data will be low.[94]

Data is also viewed as an experience good from the perspective of undertakings, which means that its value is only attained after it has been put to use and have no intrinsic value of its own.[95] In other words, data is a good that its "consumers" (i.e. undertakings) must experience to value it. As Shapiro and Varian indicated, "virtually any new product is an experience good"; however, "information is an experience good every time it's consumed."[96] The problem with experience goods is that it is difficult to determine their parameters, such as quality or price, in advance and they can be appropriately evaluated only after they have been purchased and experienced. As indicated by the OECD, estimating the monetary value of data (in particular personal data) *ex ante* (before its use) is hardly possible since it can be used in many different situations for numerous purposes and its value depends on the intended use.[97] Moreover, a particular dataset may be valuable for some applications, but useless for others.

Furthermore, another feature that distinguishes data from other assets is that its use is characterised by increasing returns to scale and scope.[98] In case of data, increasing returns to scale is reflected by the fact that data can be used to improve data-driven services, which, in turn, can attract more users. As a result, the latter generates even more data that can be collected (it is the so-called positive feedback loop).[99] For example, the more users use a search engine, the more feedback it receives and the better is the search engine's ability to return results corresponding to what users are searching for, which, in turn, reinforces the market power of the search engine.[100] As indicated by Shapiro and Varian, this "positive feedback makes the strong get stronger and the weak get weaker,

[93] See: Tucker and Welford, "Big Mistakes," 3.

[94] Acquiring personal data from data brokers is significantly limited in the EU due to restrictive EU data protection regulations applicable to processing of personal data. Further information in that regard is provided in section 2.3. below.

[95] OECD, *Data-driven Innovation*, 197; GSMA, "The Data Value Chain," 9.

[96] Shapiro and Varian, "Information Rules," 5.

[97] OECD, "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value," OECD Digital Economy Papers, No. 220, DSTI/ICCP/IE/REG(2011)2/FINAL, 4–6.

[98] OECD, *Data-driven Innovation*, 184–185; Bertin Martens, "An economic policy perspective on online platforms. JRC/IPTS Digital Economy Working Paper 2016-05," 36, available at https://ec.europa.eu/jrc/sites/default/files/JRC101501.pdf.

[99] OECD, *Data-driven Innovation*, 184. Bourreau, de Streel and Graef, "Big Data and Competition Policy," 33–34.

[100] European Commission, "Enter the data economy, EU Policies for a thriving data ecosystem, 2017 EPSC Strategic notes," 6.