Defining Big Data

Andrea De Mauro

Department of Enterprise Engineering
University of Rome Tor Vergata
Via del Politecnico, 1, 00133, Rome, Italy
E-mail: andrea.de.mauro@uniroma2.it

To cite this article: *De Mauro, A. (2023). Defining Big Data. In: Montag, C., Baumeister, H. (eds) Digital Phenotyping and Mobile Sensing. Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Cham. https://doi.org/10.1007/978-3-030-98546-2_26.*

Abstract: 'Big Data' has shown to be a trending catchphrase in modern Information Technology. However, it is hard to define a sharp cut between the classic notion of data and the novelties introduced by the arrival of Big Data. This short paper defines the seven structural elements underlying the concept of Big Data, highlighting the features that make it truly different from conventional data.

'Big Data' has shown to be a trending catchphrase that expresses the most promising trends in modern Information Technology. As of 2011, the term has increasingly permeated business readings and scholarly literature, quickly reaching the infamous status of a "hype" (Gandomi and Haider 2014, Vassakis et al. 2018). Nowadays, its current and potential applications spare no industry or research domain, impacting business and society unprecedentedly (Mayer-Schönberger and Cukier 2013). However, besides its extensive popularity, there is no complete consensus on its true definition, and it is easy to find different and diverging senses of the same term (Mikalef et al. 2017). This short paper defines the structural elements underlying the concept of Big Data, highlighting the features that make it different from conventional data.

Since the dawn of humanity, data has been the fundamental tool for recording and managing business exchanges, making it ubiquitous in any economic activity. As a consequence, firms have always leveraged data to support the functioning of their operating model. Therefore, it is hard to define a sharp cut between the traditional concept of data and the current utilization of Big Data. The literal meaning of the adjective 'Big' has led many to interpret the volume of data as its primary defining factor. However, given the continuous exponential growth of the world's technological capacity to store, communicate, and compute digital information (Hilbert and López 2011), one cannot set a static threshold of data volume. Hence, we shall ask ourselves, what makes Big Data 'Big,' and what set of

features differentiate it versus the traditional data that has been consistently used to record economic transactions.

With the emergence of the Big Data phenomenon, some of the aspects describing the characteristics, the origin, the usage, and the role of data in firms have progressively become more prominent. Table I shows seven defining data features that collectively illustrate the fundamental changes developed as the Big Data phenomenon appeared.

Defining features	Pre-Big Data	Post-Big Data
Data Volume	Stored locally or in centralized architectures	Disseminated in a remote and decentralized infrastructure (cloud)
Data Velocity	Follow times of human transactions. Data usually change over days, weeks, months	Change faster than usual human transactions, often in real-time
Data Variety	Mainly structured data, organized in tables within relational databases	More unstructured data, including natural language, images, and videos
Analytical Methods	Focus on summary statistics, aggregations, descriptive analytics	Focus on predictive and prescriptive analytics, powered by artificial intelligence
Source	Primarily from centralized entities such as firms, governments, science laboratories	Primarily from decentralized entities such as individuals and IoT devices
Primary Business Usage	Supporting accounting, controlling, and human-driven decision making	Enabling machine-driven processes and automation of cognitive tasks
Primary Strategic Role in Firms	Supports the firms' operating model: is a driver of efficiency	Participates in the firms' business model: is a strategic source of competitive advantage

Table I: Defining features of data in the Pre- and Post-Big Data era.

Volume, Velocity, and Variety (generally referenced as the "three Vs" model) have been the usual trajectories used to explain the essential characteristics of Big Data assets. In fact, the extensive size, speed of change, and increasing unstructuredness of data sets require specific technology (such as cloud-based decentralized infrastructures) and analytical methods (primarily powered by artificial intelligence) for their transformation into value (De Mauro et al. 2016). Notably, size by itself is not sufficient to explain the impact of Big Data in enhancing firm performance (Ghasemaghaei and Calic 2020). In contrast, it certainly defines the storage and computing performance requirements of the supporting technology stack.

Another defining feature of Big Data is its sources: while in the past data was mainly generated by businesses themselves and other centralized entities, data originates today from a plethora of decentralized sources. Individual consumers generate an ever-increasing feed of data points through their continuous interaction with both the digital and the physical world, forcing companies to thoroughly rethink their ways of interacting with end-consumers and understanding their behavior (Erevelles et al. 2016). The extensive availability of sensor-enabled, connected objects (known as the Internet of Things, IoT) is an additional source of distributed information that accompanied the arrival of the Big Data phenomenon (Atzori et al. 2010, Marjani et al. 2017).

Lastly, Big Data has proven to intrinsically change the way companies compete, consistently enhancing firm results when leveraged (Wamba et al. 2017, Ferraris et al. 2019). Especially when leveraged in conjunction with Artificial Intelligence, data has moved from having a supporting role (enabling accounting and operational effectiveness) to becoming a strategic source of competitive advantage and a core element of a firm's business model (Iansiti and Lakhani 2020, Sestino and De Mauro 2021).

The seven features presented in this paper provide a joint description of the novelties brought by the arrival of Big Data and explain the fundamental shifts observed versus the traditional utilization of data in firms. Interestingly, such transitions do not appear to have stopped and can be used in further research to anticipate the directional evolution of the role of data in business.

References

Atzori L, Iera A, Morabito G (2010) The Internet of Things: A survey. Comput Networks 54:2787–2805.

De Mauro A, Greco M, Grimaldi M (2016) A formal definition of Big Data based on its essential features. Libr Rev 65:122–135. https://doi.org/10.1108/LR-06-2015-0061

Erevelles S, Fukawa N, Swayne L (2016) Big Data consumer analytics and the transformation of marketing. J Bus Res 69:897–904. https://doi.org/10.1016/j.jbusres.2015.07.001

Ferraris A, Mazzoleni A, Devalle A, Couturier J (2019) Big data analytics capabilities and knowledge management: impact on firm performance. Manag Decis 57:1923–1936. https://doi.org/10.1108/MD-07-2018-0825

Gandomi A, Haider M (2014) Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage 35:137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Ghasemaghaei M, Calic G (2020) Assessing the impact of big data on firm innovation performance: Big data is not always better data. J Bus Res 108:147–162. https://doi.org/10.1016/j.jbusres.2019.09.062

Hilbert M, López P (2011) The world's technological capacity to store, communicate, and compute information. Science 332:60–65.

Iansiti M, Lakhani KR (2020) Competing in the age of AI: strategy and leadership when algorithms and networks run the world. Harvard Business Press

Marjani M, Nasaruddin F, Gani A, Karim A, Hashem I, Targio A, Siddiqa A, Yaqoob I (2017) Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. IEEE Access 5:5247–5261. https://doi.org/10.1109/ACCESS.2017.2689040

Mayer-Schönberger V, Cukier K (2013) Big Data: A Revolution That Will Transform How We Live.

Mikalef P, Pappas IO, Krogstie J, Giannakos M (2017) Big data analytics capabilities: a systematic literature review and research agenda. Inf Syst E-bus Manag 1–32. https://doi.org/10.1007/s10257-017-0362-y

Sestino A, De Mauro A (2021) Leveraging Artificial Intelligence in Business: Implications, Applications and Methods. Technol Anal Strateg Manag 1–14. https://doi.org/10.1080/09537325.2021.1883583

Vassakis K, Petrakis E, Kopanakis I (2018) Big Data Analytics: Applications, Prospects and Challenges. In: Mobile Big Data. pp 3–20

Wamba SF, Gunasekaran A, Akter S, Ren SJ, Dubey R, Childe SJ (2017) Big data analytics and firm performance: Effects of dynamic capabilities. J Bus Res 70:356–365. https://doi.org/10.1016/j.jbusres.2016.08.009