Perspectives to Definition of Big Data: A Mapping Study and Discussion

Ossi Ylijoki[1*], Jari Porras[1]

[1]School of Business and Management, Lappeenranta University of Technology, Finland
`ossi.ylijoki@phnet.fi, jari.porras@lut.fi`

Abstract. Big data is an emerging research area where common terminology is still evolving. Different perspectives to the research area and terminology exist, but a common definition for big data does not exist. We have performed a systematic mapping study in order to identify different big data definitions and their perspectives. As a result, we present a state-of-the-art review of the current status in big data definitions, discuss the shortcomings of the current definitions, and propose possible solutions for the shortcomings. The paper contributes to the emerging big data research by analyzing current definitions of big data from different perspectives, suggesting enhancement to the terminology as well as pointing out new research avenues. In addition, the article helps new researchers and practitioners to understand what big data is, and bridges the knowledge between theory and practice.

Keywords. Analytics, Big Data, Big Data Definition, Business Model, Datafication, Digitization, Knowledge Management.

1   Introduction

Digitization is a current megatrend, meaning that digital technologies are integrated into our everyday life. The use of digital technologies enables the connection of different services and automation of many processes. Although digitization itself is an important technological (r)evolution, it enables even more fundamental change: datafication. An increasing number of devices and sensors are constantly connected to the Internet. Cameras, mobile phones, tablets, various applications and services running on them produce wide varieties of digital data. This data generation phenomenon is called datafication (Mayer-Schönberger and Cukier, 2013). Lycett (2013) defines datafication as a "sense-making process", which emphasizes the value generation aspect. Digitization and datafication make it possible to capture different situations, actions, or even series of events in the form of data. A vague term "big data" describes the data resulting from datafication. This phenomenon has widespread effects.

As an example, let us consider quadcopters. Amazon and DHL[1], among others, are prototyping these small flying devices for delivering goods to customers. Quadcopters

---

1   Amazon Prime Air. http://www.amazon.com/b?node=8037720011. Accessed 28th April 2016.
   DHL: http://www.theguardian.com/technology/2014/sep/25/german-dhl-launches-first-commercial-drone-delivery-service. Accessed 28th April 2016.

gather vast volumes of different types of data in real-time (e.g. sensor readings, video and geolocation data) in order to be able to perform their tasks autonomously. They analyze and use the data in many tasks, such as avoiding collisions and orienting their way to the destination. In addition, they synthetize and distribute data. Sending data, such as location and altitude to the command center is essential for the fleet management.

As the fleet of a firm might contain thousands of quadcopters, this represents a real-world big data problem. In general, there are numerous *technical challenges* to conquer for organizations that wish to benefit of big data, see e.g. (Ma et al., 2013; Chen et al., 2014; Kambatla et al., 2014). So far, humans supervise most quadcopter experiments, but due to rapid technical advances, it is obvious that in the near future these little flying machines will become autonomous. Hardware and software vendors are investing heavily in their offerings, so this area is progressing rapidly.

Big data resulting from digitization is seen as a *significant opportunity*, see e.g. (Manyika et al., 2011; Mayer-Schönberger and Cukier, 2013; Schmarzo, 2013; Davenport, 2014). Big data is considered as a key enabler that can be used to generate value in private companies and public organizations. Governments have initiated big data strategies[2]. Examples of the benefits include creating new business opportunities, boosting R&D activities, and supporting decision making (Amatriain, 2013; Mehta et al., 2013; Lee et al., 2014). Quadcopters, among other technological solutions, can be used to save costs and even enable new business models for many organizations, both in the private and the public sector. There are naturally also questioning voices that criticize the big data paradigm and value proposals (see e.g. (boyd and Crawford, 2012; Fox and Do, 2013).

Datafication and big data are disruptive technologies that have widespread implications on the society. Technology vendors, the public sector, private companies, consumers, and policy makers, among others, have interests in the field. Moreover, as the number of stakeholders and parties increases, common understanding of the terminology and concepts becomes more and more important. Unfortunately, big data is a volatile term now. Different definitions of big data can be found in the literature, as well as among practitioners. A (theoretical) definition is a proposal for understanding the meaning of a term. It should be observable, clear (i.e., unambiguous) and complete. Good definitions improve the quality of communication significantly and enable common understanding among participants from different backgrounds. To put it simply, a good definition equals clarity.

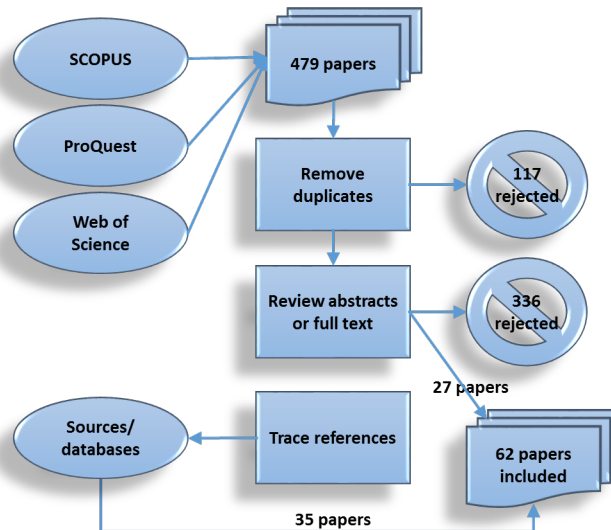The purpose of this article is – considering the broad implications of big data on the

---

2 E.g. European Big Data Value Strategic Research & Innovation Agenda.
   http://www.nessi-europe.eu/Files/Private/EuropeanBigDataValuePartnership_SRIA__v099%20v4.pdf.
   Accessed 28th April 2016.
   U.S. Big data initiative:
   https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf. Accessed 28th April 2016.

society, organizations, and individuals – to shed light on the definition of big data. As the method, we use a systematic mapping study. According to Kitchenham (2007), mapping studies are designed to give a broad overview of a research area. Mapping studies have typically broad research questions. Our research questions are:

- What kind of definitions of big data exist in research papers and among practitioners?
- How has the definition of big data evolved?
- How do the definitions reflect the different characteristics and perspectives of big data?

2   Literature Search

Our initial search covered three major reference databases: Scopus, ProQuest, and Web of Science. We considered this as a good starting point, as these databases index a broad range of papers, covering both technical and business fields. Figure 1 gives an overview of the search process. In addition to wide research questions, Kitchenham (2007) suggests that mapping studies should use rather loose search criteria. We searched the databases (title, keywords, abstract) by using ("big data" and "definition") as a search string. All papers written in English and indexed up to 02-Sep-2015 were included in the initial result set. No additional limitations were set. A total of 479 papers were identified. Next, we removed duplicate articles (117).

Fig. 1. Search process.

After removing the duplicates, we read the abstracts, and where necessary, the whole text of each of the resulting papers. We categorized the papers by using the following inclusion/exclusion criteria: If the paper contains a definition of big data, include it, otherwise reject it. Due to the loose search criteria, a number of papers defining other

things than big data were included in the initial search. Papers that obviously did not meet the eligibility criteria were rejected. If the decision was not clear, we performed a full text review, and the paper was either included or excluded on the basis of the review. Additional 17 papers were excluded because they were either commercial, high-price reports or they could not be found. As a result of this phase, 27 papers were included in the result set.

In the reference-tracking phase, we searched for additional papers on the basis of citations in the included papers (backward snowballing). Possibly interesting references were checked in the article context, and if still promising, they were tracked from databases or web sources, including Google Scholar and various web pages. If the article met the eligibility criteria, it was included. We identified additional 35 papers in this phase.

At the end of the search process phase, we had identified 62 papers that contained a definition of big data. The year-wise distribution of these papers is presented in Figure 2. It seems that although the first definition was presented more than 10 years ago, the discussion of the definition of big data started only a few years back. These papers and their definitions were examined further.
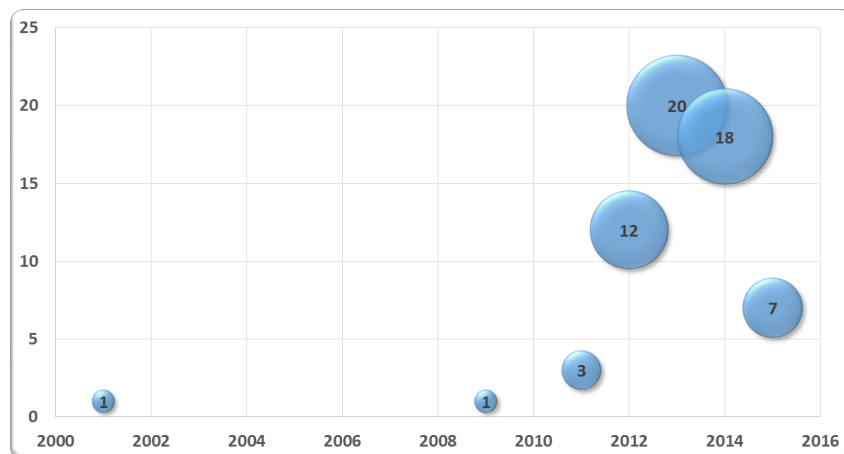
Fig. 2. Year-wise distribution of found papers.

3   Analysis of the Definitions

The first part of the analysis covers the evolution of the definition of big data. Definitions, their existence in time, as well as similarities and differences are presented. This analysis reveals what perspectives (or components) various participants have added to the definition over time. The second part of our analysis identifies gaps between the current definitions and big data value propositions, in order to find out what perspectives are still missing.

3.1 Evolution of the Term Big Data

The term "big data" is not new. It has been used both in research and non-research papers for quite a long time. Back in 1997 it was used in the context of visualizing large data sets (Cox and Ellsworth, 1997). In 1998 it was used in a hardware-related presentation (Mashey, 1998) and also in the data mining context (Weiss and Indurkhya, 1998), and 2003 in combination with statistics (Diebold, 2003). In the beginning, big meant the size and all these sources recognized and referenced big data with the increasing volumes of data. However, year 2001 can be considered as a major milestone in the definition of big data. Laney (2001) described three essential dimensions of big data: *volume*, *velocity* and *variety*. Operating with a swarm of autonomous quadcopters requires the management of high-volume, high-velocity (real-time) data that have many types (variety).

During the following decade, trailblazers like Google and Amazon developed practical big data solutions. These solutions have proved to add value to their businesses. In fact, the trailblazers build their business models on big data solutions. An article published in 2008 in the Wired magazine (Anderson, 2008) aroused public interest in the use of big data and its effects in science. The next significant milestone was 2011, when McKinsey Global Institute and IDC published reports (Gantz and Reinsel, 2011; Manyika et al., 2011) that drew wide public attention to the potential value of big data. Since then a number of newspaper articles, scientific big data papers and books have been published.

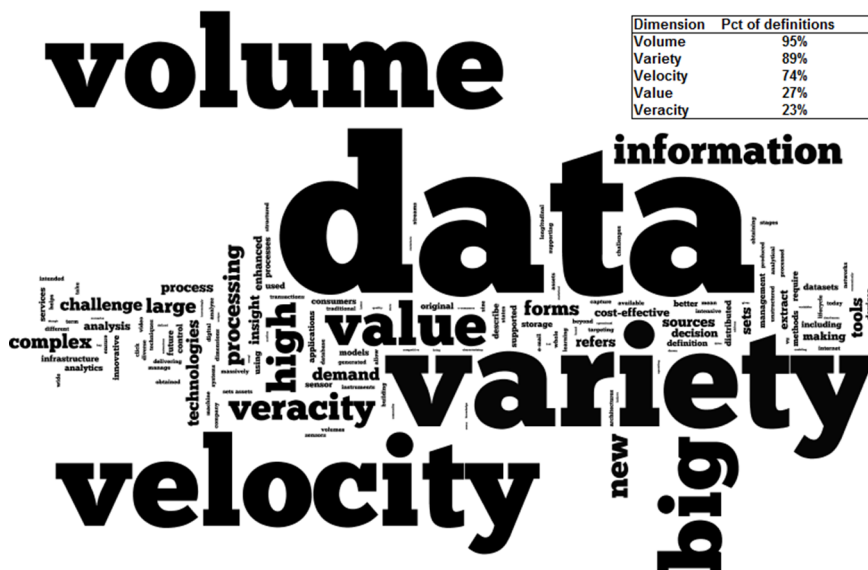| Dimension	Pct of definitions
Volume	95%
Variety	89%
Velocity	74%
Value	27%
Veracity	23%

Fig. 3. Most common big data characteristics.

We considered Laney (2001) to be the one to offer the first real definition, although the term big data had been used earlier. In our analysis of the studies, we could not identify references to earlier papers. Naturally, Laney must have been influenced by

earlier work, but his paper was the first to introduce the three big data dimensions: volume, variety and velocity. Most of the definitions rely at least partly on the 3V definition by Laney (2001). Figure 3 shows the most common characteristics used in the definitions of big data (see Appendix 1 for details of the definitions). 95% (59 occurrences out of 62) of the papers identified volume as a key characteristic of big data. In addition, the papers considered variety (55 occurrences) and velocity (46) to be typical big data factors. Value (17) and veracity (14) had also caught attention. These five dimensions dominate the current definitions of big data.

The included 62 papers (see Appendix 1 for details) were arranged by their publishing date, and each paper was inspected against previously published definitions. If the paper contained a new definition or added some new elements to the existing definitions, it was considered to be a new definition. This analysis resulted in 17 different definitions. These 17 definitions have similarities in the sense that many of them aim to widen the 3V definition to cover technical and especially business aspects. This is quite a natural consequence with regard to the big data value proposal. However, wide definitions can be problematic, and some essential aspects of big data are still lacking. We will discuss these aspects below. The rest of the papers (45) contained definitions essentially covered in earlier papers. Appendix 1 presents details of the definitions.
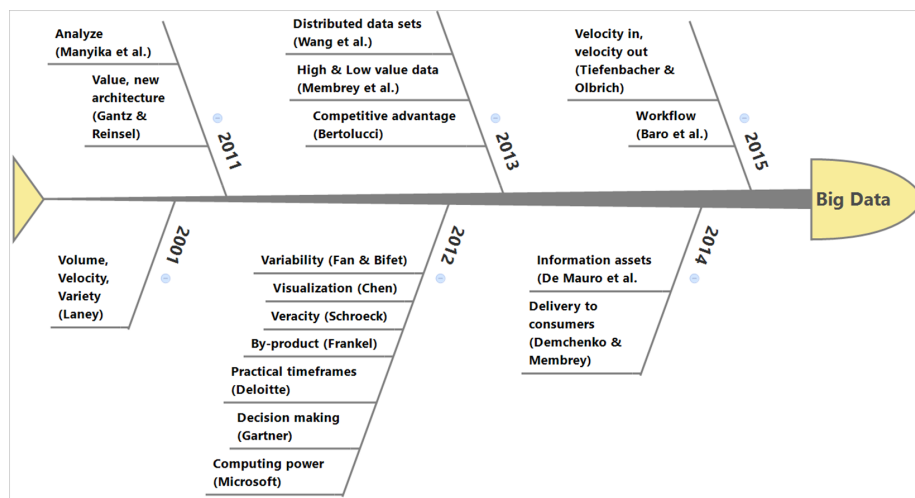

Fig. 4. Evolution of the definition of big data.

The fishbone diagram in figure 4 gives an overview of the evolution. The bones show essential additions of all 17 different definitions, i.e. new aspects or components that each definition adds. Laney (2001) presented the original, so-called 3V definition of big data. The Vs come from volume, velocity and variety. Volume refers to ever-increasing amounts of data. Velocity indicates the need to capture and analyze high-speed or bursts of data in (near) real-time, or else the value may be lost. Variety is related to different types of data, be it structured or non-structured, such as social media posts or a video.

The 3V definition was the de-facto big data standard until 2011, when both Manyika et al. (2011) and Gantz and Reinsel (2011) published their reports. Manyika et al. (2011) emphasize the potential value of big data, but curiously enough, their definition focuses on data volume including only a hint ("analyze") of the value. Also, when compared to Laney (2001), Manyika et al. (2011) have left out velocity and variety. Gantz and Reinsel (2011) include the three Vs, and add value extraction and new architectures. They have also decided to define big data technologies instead of big data. This approach allows them to balance the definition between data, technology and business components with a reasonable logic.

The big data hype was at its peak in the years 2012 and 2013. Several aspects of big data were discussed, such as privacy, security, (business) value, and veracity. We identified seven definitions from 2012 that were either completely new, like the one by Microsoft (2012), or added new components to existing definitions (Gartner, 2012; Schroeck et al., 2012; Fan and Bifet, 2013), and three from the year 2013. After that date we identified four more additions. Two of these later definitions (Demchenko, DeLaat, et al., 2014; Baro et al., 2015) note the importance of delivering the results to consumers. This analysis showed that the evolution of the definition started with data and especially data volumes, and then the discussion shifted to infrastructure topics, followed by the (business) value of data. Finally, more fine-grained aspects, like data delivery and collaboration, appeared.

3.2   Definitions vs. Big Data Value Chain

An interesting question is how the 17 different big data definitions reflect the significant value proposal of big data? Several frameworks explain how data adds value. One of the first of such models is the Virtual Value Creation (VVC) framework presented by Rayport and Sviokla (1995). This framework describes five steps that are required to create value from data: gather, organize, select, synthesize, and distribute (see Figure 5). The steps gather and organize are data-related, and they cover aspects like data acquisition from sensors, integration with other data, and data storing. The steps select, synthesize and distribution depend on data usage. They are activities like filtering data for analysis, or represented as artifacts like analytical models, data visualization, and information delivery tools. Value is expected to increase as data items from various sources are combined to form meaningful information chunks in the VVC process.

A quadcopter reads its current location from the GPS sensor and combines it with the destination information (gather, organize). Based on the analysis, it may take a decision to change its direction (select, synthetize). At frequent intervals, the copter sends data (e.g. location) to the command center (distribute). This simple VVC process adds value, as it enables the copter to work autonomously. However, taking a helicopter view by looking at the whole fleet instead of one quadcopter, it becomes clear that much more value is available. The command center systems gathers data from each of the copters and other sources, e.g. from delivery orders (gather, organize). An analytical model calculates the routes (select, synthetize) and sends instructions (like pick-up and delivery addresses) to each of the copters (delivery). This automated VVC process creates value from the data by producing optimal routes, maximizing the number of deliveries and increasing efficiency.
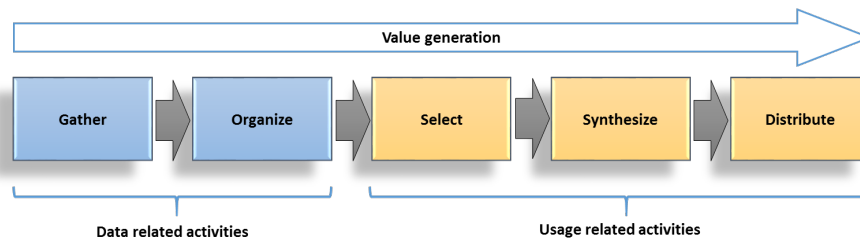
Fig. 5. Virtual value creation process.

Table 1 maps the 17 different definitions to the big data value chain. Together the current big data definitions cover all phases of the value chain. However, most of the definitions cover only parts of the chain. There are two definitions that consider all five phases, those of Demchenko, DeLaat, et al. (2014) and Baro et al. (2015).

Note that the table shows which phases of the value chain the new perspective of each definition emphasizes. This is for clarity: many of the definitions cover also other phases, e.g. Demchenko, DeLaat, et al. (2014) have also covered other steps. However, the new perspective of their definition is the delivery aspect, and therefore only the distribute phase is included in Table 1.

Table 1. Mapping the new perspectives of the big data definitions to the value chain.

Authors	New perspective	Gather (data)	Organize (data)	Select (usage)	Synthesize (usage)	Distribute (usage)
(Laney, 2001)	Volume, velocity, variety	x	x			
(Manyika et al., 2011)	Analyze			x	x	
(Gantz and Reinsel, 2011)	Value, new architecture		x	x	x	
(Microsoft, 2012)	Computing power		x	x	x	
(Gartner, 2012)	Decision making			x	x	
(DeloitteConsulting, 2012)	Practical timeframes		x	x	x	
(Frankel, 2012)	By-product	x				
(Schroeck et al., 2012)	Veracity	x	x	x	x	
(Chen et al., 2012)	Visualization				x	
(Fan and Bifet, 2013)	Variability	x				
(Wang et al., 2013)	Distributed data sets			x		
(Membrey et al., 2013)	High & low value data			x	x	
(Bertolucci, 2013)	Competitive advantage				x	
(Demchenko, DeLaat, et al., 2014)	Delivery to consumers					x
(De Mauro et al., 2015)	Information assets			x		
(Tiefenbacher and Olbrich, 2015)	Velocity in, velocity out	x	x	x	x	
(Baro et al., 2015)	Workflow					x

4   Discussion

As can be seen in the definitions and analysis, big data can mean different things, depending on the selected viewpoint. Some perceive big data as a technical challenge, others view it as a vehicle to increase efficiency or profits. In this section we will show that combining data and its intended usage leads to vague definitions, and consider how the disruptive nature of big data should be taken into account.

4.1   Separate Data and Its Usage

Our analysis revealed that several definitions have logical incoherencies. Value, for example, must be derived from the data by using analytics, there is no value in plain data as such (Ackoff, 1989). Value is also case-dependent. A certain piece of information may be worthless to one company but highly valued by some other firm or in another situation. For example, quadcopter flight details are much more valuable in case of an accident than in a normal situation. This is emphasized by Mayer-Schönberger and Cukier (2013) who state that the value of big data is in the secondary uses of the data. For veracity, analytics is required to determine whether the data is relevant for the planned usage. As important factors as value and veracity are in practice, they do not define the characteristics of big data, but instead they reflect the *usage* of the data. Vague definitions are typically hard to understand as they raise questions that cannot be answered coherently. This will lead to different interpretations and misunderstandings.

The original 3V definition (Laney, 2001) left the business effects out. This is one of the main reasons why many new definitions have emerged. Both technology vendors and enterprises have an interest to add a value proposition. Companies see big data as a vehicle to gain value, vendors naturally like to justify the costs of their offerings with potential benefits. A natural tendency would be to add a value component to the definition. However, as discussed above, value is not a characteristic of data. Definitions should be clear and unambiguous. Therefore, adding data usage to the definition is not a good idea, as the definition would become unambiguous, and coherency would be lost.

Our suggestion to the problem is that the *data and its usage should be separated*. Data is similar to oil: when combined with data management and analytics processes it provides organizations with value. Analytics and data usage are of course essential elements in successful big data exploitation. However, from the definition point of view, combining data and its usage is like combining oil and engine into one single definition. Separating big data from its intended usage clarifies the inconsistencies of the definitions and helps us to understand the plain characteristics of big data. As the purpose of data usage is to realize the potential value of the data, we propose the term *big data insights* to be used in any context in data usage -related activities (see also figure 5).

4.2   Other Perspectives – Big Data as a Phenomenon

In addition to technical and value aspects, scholars have focused on several other perspectives to big data, such as privacy, security (Altshuler, 2011; Berghel, 2013; Lu et al., 2014), and policy-making (Keen et al., 2013; Blume et al., 2014; Truyens and Van Eecke, 2014). None of the current definitions of big data consider these. These aspects are not characteristics of big data; we do not suggest that these aspects should be included in the definition. Instead, they are aspects that help to understand big data as a phenomenon. Moreover, these perspectives are important, as failing to consider them can drive an organization to difficulties.

Another, even more important aspect is that the current definitions neglect the disruptive nature of big data. On the basis of the literature it seems obvious that in the future, big data will have significant impacts on businesses (Manyika et al., 2011; Schmarzo, 2013; Davenport, 2014). Big data is seen as a technology that can have huge impacts on most industries and enterprises. Data-driven companies can achieve significant benefits (McAfee et al., 2012), but transformational business changes (Dehning et al., 2003) are required to achieve full competitive advantage from big data. The impact of big data will be significant, but the nature of the change is even more important. The effects of big data on firms, ecosystems and industries will be disruptive (Earley, 2014; Fan and Gordon, 2014; Kim et al., 2014). Industry structures are changing, and new business opportunities are emerging. On the other hand, this means that also competitors may be able to invent new business models, not to speak of new entrants, which will increase the turbulence effectively. The impacts of big data may – and will – be positive for some organizations, negative for others. Due to the disruptive nature of big data, companies must review their business models in order to reveal possible threats and opportunities. Moreover, as the disruptive drivers are technological by nature, these technologies and their potential effects must be linked with strategy.

We suggest that a *new definition for big data as a phenomenon should be considered*. For clarity and coherency, the definition of big data should cover only data and data management aspects (like the 3V definition). The phenomenon of big data is a broad concept that deserves a definition of its own. Instead of defining big data, the definition should consider several important perspectives of it. In our opinion, this definition should include the disruptive nature and strategic importance of the phenomenon. Adding these elements would help managers to understand the importance of the matter. This opens a new research avenue. Discussing and defining the nature and relations between various perspectives would build understanding of the broader context of big data, big data as a phenomenon.

5   Conclusions

Our aim was to shed light to the concept of big data, especially from the following viewpoints:

- What kind of definitions of big data exist in research papers and among practitioners?

- How has the definition of big data evolved?
- How do the definitions reflect the different characteristics and perspectives of big data?

A systematic mapping study was conducted in order to find answers to these questions. We made a search in major reference databases, search engines, and web sources containing both technical and business topics. A total of 62 sources were included in the result set. With regard to our research questions, we chose a broad search strategy in order to cover a wide range of possible sources. We identified 17 different definitions of big data that together presented a clear picture of the current situation and evolution of the definition, thus providing answers to our first and second research questions. We also compared the current definitions with various characteristics of big data. We found that the current definitions do not cover several perspectives that are discussed among big data scholars and practitioners, which answers our third research question. In addition, we identified several logically incoherent definitions. This clouds the matter further, as these definitions raise new questions, which will typically lead to ambiguous answers.

5.1   Results

This study revealed 17 different big data definitions from 62 relevant source papers. Each of the papers was analyzed against previously published definitions. If the paper contained a new definition or added some new elements to the existing definitions, it was considered as a new definition. The key contributions of this study are:

- Although there are various opinions on what big data is, the 3V definition by Laney (2001) contains three dimensions (volume, velocity, variety), which are common to most definitions. In addition to these dimensions, many definitions include technical parts and components related to the intended usage of the data, such as analysis or decision-making.
- Many of the definitions are logically inconsistent, which is one reason for the vagueness of the term big data. A typical flaw is to include both the data and its intended usage in the definition. We suggest that they should be separated. The term big data should cover data-related aspects, whereas a new term *big data insights* should be used when discussing data usage-related activities.
- The current definitions do not consider several important aspects of the big data phenomenon, such as security and privacy, or its disruptive nature. These are not characteristics of big data, but they are important factors of the big data phenomenon that both scholars and practitioners must consider. We suggest that a new definition for big data as a phenomenon should be developed.

In addition, this study bridges the knowledge between theory and practice. We have presented the history and the state-of-the-art of the definition of big data. This will help new researchers and practitioners to understand the different perspectives of big data, as well as the limitations of the current definitions. Therefore, we hope that this paper will also stimulate discussion about the terminology and help parties coming from different backgrounds to understand each other and communicate their

reasoning clearly.

5.2   Limitations

We recognize that an uncountable number of various definitions of big data exist in the "Internet jungle", e.g. in blog postings and discussion forums. However, due to limited resources, identifying and analyzing all or even most of them would be impossible, and therefore we have filtered blogs and forums out. Another limitation is that we have excluded all non-English language sources.

5.3   Suggestions for Further Studies

There are several possible topics for further studies, including the following. It is clear that there is a need to develop the terminology and taxonomy further (including related terms, such as big data analytics, big data phenomenon, and veracity) in order to create common understanding of the key concepts and their relationships in the area of big data. Another interesting research avenue would be to investigate the effects of big data on organizations' business models or decision-making processes, organizational structures, and culture.

6   References

Ackoff, R.L., 1989. From Data to Wisdom. Journal of Applied Systems Analysis 16, 3–9.

Akerkar, R., 2014. Analytics on Big Aviation Data: Turning Data Into Insights. International Journal of Computer Science and Applications 11, 116–127.

Altshuler, M. Y.; Aharony N.; Pentland A.; Elovici Y.; Cebrian, 2011. Stealing Reality: When Criminals Become Data Scientists (or Vice Versa). IEEE Intelligent Systems 26, 22–30.

Amatriain, X., 2013. Beyond data: from user information to business value through personalized recommendations and consumer science. Proceedings of the 22nd ACM international conference on information & knowledge management 2201–2208.

Anderson, C., 2008. The end of theory: The data deluge makes the scientific method obsolete. Wired Magazine, 23-Jun-2008.

Ashraf, J., Hussain, O.K., Hussain, F.K., 2015. Making sense from big RDF data: OUSAF for measuring ontology usage. Software: Practice and Experience 45, 1051–1071.

Balar, A., Malviya, N., Prasad, S., Gangurde, A., 2013. Forecasting consumer behavior with innovative value proposition for organizations using big data analytics. Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on 1–4.

Baro, E., Degoul, S., Beuscart, R., Chazard, E., 2015. Toward a Literature-Driven Definition of Big Data in Healthcare. BioMed Research International 2015, 1–9.

Benjamins, V.R., 2014. Big Data: from Hype to Reality? Proceedings of the 4th

International Conference on Web Intelligence, Mining and Semantics (WIMS14) 2.

Berghel, H., 2013. Through the PRISM darkly. Computer 46, 86–90.

Bertolucci, J., 2013. Big Data: A Practical Definition. Informationweek - Online.

Blume, G., Scott, T., Pirog, M., 2014. Empirical Innovations in Policy Analysis. Policy Studies Journal 42, S33–S50.

boyd, danah, Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society 15, 662–679.

Cackett, D., 2013. Information Management and Big Data: A Reference Architecture. Oracle Corporation. Accessed 28th April 2016. http://www.oracle.com/technetwork/topics/entarch/articles/info-mgmt-big-data-ref-arch-1902853.pdf.

Chen, H., Chiang, R.H., Storey, V.C., 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly 36, 1165–1188.

Chen, M., Mao, S., Liu, Y., 2014. Big data: A survey. Mobile Networks and Applications 19, 171–209.

Collins, E., 2014. Intersection of the cloud and big data. IEEE Cloud Computing 1, 84–85.

Cox, M., Ellsworth, D., 1997. Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th conference on Visualization'97 235–ff.

Cuzzocrea, A., Song, I.-Y., Davis, K.C., 2011. Analytics over large-scale multidimensional data: the big data revolution! Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP 101–104.

Davenport, T., 2014. Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press.

De Mauro, A., Greco, M., Grimaldi, M., 2015. What is big data? A consensual definition and a review of key research topics. AIP Conference Proceedings 1644, 97–104.

Dehning, B., Richardson, V.J., Zmud, R.W., 2003. The value relevance of announcements of transformational information technology investments. Mis Quarterly 27, 637–656.

DeloitteConsulting, 2012. The insight economy: Big data matters—except when it doesn't. Deloitte. Accessed 28th April 2016. http://www2.deloitte.com/content/dam/Deloitte/global/Documents/Deloitte-Analytics/dttl-analytics-us-ba-insight-economy-10012012.pdf.

Demchenko, Y., DeLaat, C., Membrey, P., 2014. Defining architecture components of the Big Data Ecosystem. Collaboration Technologies and Systems (CTS), 2014 International Conference on 104–112.

Demchenko, Y., Grosso, P., De Laat, C., Membrey, P., 2013. Addressing big data issues in scientific data infrastructure. Collaboration Technologies and Systems (CTS), 2013 International Conference on 48–55.

Demchenko, Y., Gruengard, E., Klous, S., 2014. Instructional Model for Building Effective Big Data Curricula for Online and Campus Education. Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on 935–941.

Demchenko, Y., Ngo, C., Laat, C. de, Membrey, P., Gordijenko, D., 2014. Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure, in: SecureDataManagement. Springer, pp. 76–94.

Diebold, F.X., 2003. Big Data'Dynamic factor models for macroeconomic measurement and forecasting, in: Advances EconomicsEconometricsTheory Applications, Eighth World CongressEconometricSociety(edited M. Dewatripont, LP HansenS.Turnovsky). pp. 115–122.

Earley, S., 2014. The Digital Transformation: Staying Competitive. IT Professional 16, 58–60.

Emani, C.K., Cullot, N., Nicolle, C., 2015. Understandable Big Data: A survey. Computer Science Review 17, 70–81.

EMC, 2012. Big Data-as-a-Service. EMC Solutions Group. Accessed 28th April 2016. http://www.emc.com/collateral/software/white-papers/h10839-big-data-as-a-service-perspt.pdf.

Fan, W., Bifet, A., 2013. Mining big data: current status, and forecast to the future. ACM sIGKDD Explorations Newsletter 14, 1–5.

Fan, W., Gordon, M.D., 2014. The power of social media analytics. Communications of the ACM 57, 74–81.

Ferrando-Llopis, R., Lopez-Berzosa, D., Mulligan, C., 2013. Advancing value creation and value capture in data-intensive contexts. Big Data, 2013 IEEE International Conference on 5–9.

Fox, S., Do, T., 2013. Getting real about Big Data: applying critical realism to analyse Big Data hype. International Journal of Managing Projects in Business 6, 739–760.

Frankel, D.A., 2012. Big Data and Risk Management. Risk Management.

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35, 137–144.

Gantz, J., Reinsel, D., 2011. Extracting value from chaos (No. 1142), IDC iview. IDC. Accessed 28th April 2016. http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.

Gardner, E., 2013. The HIT approach to big data. Health data management 21, 34–36.

Gartner, 2012. Gartner IT Glossary - Big Data. Accessed 28th April 2016. http://www.gartner.com/it-glossary/big-data.

Gerhardt, B., Griffin, K., Klemann, R., 2012. Unlocking value in the fragmented world of big data analytics, Cisco Internet Business Solutions Group, June. Cisco Internet Business Solutions Group (IBSG).

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U., 2015. The rise of "big data" on cloud computing: review and open research issues.

Information Systems 47, 98–115.

Hu, H., Wen, Y., Chua, T.-S., Li, X., 2014. Toward scalable systems for big data analytics: A technology tutorial. Access, IEEE 2, 652–687.

Jacobs, A., 2009. The pathologies of big data. Communications of the ACM 52, 36–44.

Jin, X., Wah, B.W., Cheng, X., Wang, Y., 2015. Significance and challenges of big data research. Big Data Research 2, 59–64.

Kambatla, K., Kollias, G., Kumar, V., Grama, A., 2014. Trends in big data analytics. Journal of Parallel and Distributed Computing 74, 2561–2573.

Keen, J., Calinescu, R., Paige, R., Rooksby, J., 2013. Big data+ politics= open data: The case of health care data in England. Policy & Internet 5, 228–243.

Kim, G.-H., Trimi, S., Chung, J.-H., 2014. Big-data applications in the government sector. Communications of the ACM 57, 78–85.

Kitchenham, B., 2007. Guidelines for performing systematic literature reviews in software engineering (No. EBSE-2007-01), Technical report, Ver. 2.3 EBSE Technical Report. EBSE. Keele University.

Lamont, J., 2012. Big data has big implications for knowledge management. KM World 21, 8–11.

Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META Group Research Note 6, 70.

Lee, J., Kao, H.-A., Yang, S., 2014. Service innovation and smart analytics for industry 4.0 and big data environment. Procedia CIRP 16, 3–8.

Li, T.Z., Wang, S.H., Ma, J., 2014. Study on Fair Definitions and Application Modes of Big Data, in: AppliedMechanics Materials. Trans Tech Publ, pp. 606–613.

Lin, P.P., 2014. What CPAs Need to Know about Big Data. The CPA Journal 84, 50.

Lu, R., Zhu, H., Liu, X., Liu, J.K., Shao, J., 2014. Toward efficient and privacy-preserving computing in big data era. Network, IEEE 28, 46–50.

Lycett, M., 2013. Datafication': Making sense of (big) data in a complex world. European Journal of Information Systems 22, 381–386.

Ma, M., Wang, P., Chu, C.-H., 2013. Data management for internet of things: challenges, approaches and opportunities. Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing 1144–1151.

Madden, S., 2012. From databases to big data. IEEE Internet Computing 16, 4–6.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Mashey, J.R., 1998. Big Data... and the Next Wave of InfraStress. Accessed 28th April 2016. http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf.

Mayer-Schönberger, V., Cukier, K., 2013. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D., 2012. Big data: The Management Revolution. Harvard Business Review 90, 61–67.

Mehta, S., Pimplikar, R., Singh, A., Varshney, L.R., Visweswariah, K., 2013. Efficient multifaceted screening of job applicants. Proceedings of the 16th International Conference on Extending Database Technology 661–671.

Membrey, P., Chan, K.C., Demchenko, Y., 2013. A disk based stream oriented approach for storing big data. Collaboration Technologies and Systems (CTS), 2013 International Conference on 56–64.

Meng, L., Meng, L., 2014. Application of Big Data in Higher Education, in: 2ndInternationalConference TeachingComputationalScience. Atlantis Press.

Microsoft, 2012. The Big Bang: How the Big Data Explosion Is Changing the World. Accessed 28th April 2016. http://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/.

O'Leary, D.E., 2013. Artificial intelligence and big data. IEEE Intelligent Systems 28, 96–99.

Pandey, S., Tokekar, V., 2014. Prominence of mapreduce in big data processing. Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on 555–560.

Pospiech, M., Felden, C., 2013. A Descriptive Big Data Model Using Grounded Theory. Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on 878–885.

Provost, F., Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. Big Data 1, 51–59.

Rayport, J.F., Sviokla, J.J., 1995. Exploiting the virtual value chain. Harvard business review 73, 75.

Richards, G., 2014. Let us clarify what we mean by Big Data; We first need a clear definition. The Ottawa Citizen, 02-Aug-2014.

Sagiroglu, S., Sinanc, D., 2013. Big data: A review. Collaboration Technologies and Systems (CTS), 2013 International Conference on 42–47.

Schmarzo, B., 2013. Big Data: Understanding how data powers big business. John Wiley & Sons.

Schneider, R.D., 2012. Hadoop for dummies, Special Edition, John Willey & sons. John Wiley & Sons Canada, Ltd.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P., 2012. Analytics: The real-world use of big data, IBM Global Business Services, Somers. IBM. Accessed 28th April 2016.
https://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf.

Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., Philippart, L., 2014. Using big data to improve customer experience and business performance. Bell Labs Technical Journal 18, 3–17.

Stonebraker, M., Robertson, J., 2013. {Big Data is'Buzzword Du Jour;'CS Academics'Have the Best Job'}. Commun. ACM 56, 10–11.

TataConsultancyServices, 2013. The Emerging Big Returns on Big Data. Tata
Consultancy Services. Accessed 28th April 2016.
http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-
Global-Trend-Study-2013.pdf.

TheIrishTimes, 2013. Get to grips with the streams of data you generate. Irish Times.

Tiefenbacher, K., Olbrich, S., 2015. Increasing the Value of Big Data Projects–
Investigation of Industrial Success Stories. System Sciences (HICSS), 2015 48th
Hawaii International Conference on 294–303.

Truyens, M., Van Eecke, P., 2014. Legal aspects of text mining. Computer Law &
Security Review 30, 153–170.

Vossen, G., 2014. Big data as the new enabler in business and other intelligence.
Vietnam Journal of Computer Science 1, 3–14.

Wang, W., Zhou, X., Zhang, B., Mu, J., 2013. Anomaly detection in big data from
UWB radars. Security and Communication Networks 8, 2469–2475.

Ward, J.S., Barker, A., 2013. Undefined by data: a survey of big data definitions.
arXiv preprint arXiv:1309.5821.

Weiss, S.M., Indurkhya, N., 1998. Predictive data mining: a practical guide. Morgan
Kaufmann.

Wyber, R., Vaillancourt, S., Perry, W., Mannava, P., Folaranmi, T., Celi, L.A., 2015.
Big data in global health: improving health in low-and middle-income countries.
Bulletin of the World Health Organization 93, 203–208.

Xin, N.Y., Ling, L.Y., 2013. How we could realize big data value. Instrumentation
and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd
International Symposium on 425–427.

Zhang, J., Chen, Y., Li, T., 2013. Opportunities of innovation under challenges of big
data. Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th
International Conference on 669–673.

Appendix 1 – Included Papers

Definitions of big data sorted by date. The Definition column contains either a direct quotation from the paper or essential parts of the given/referenced definition. The New perspective column indicates what new component or aspect the definition has added to the previous ones.

Authors	Date	Definition	New perspective
(Laney, 2001)	2001-Feb	"E-commerce, in particular, has exploded data management challenges along three dimensions: volumes, velocity, and variety."	volume, velocity, variety
(Jacobs, 2009)	2009-Aug	"data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time."	-
(Manyika et al., 2011)	2011-May	"Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."	analyze
(Gantz and Reinsel, 2011)	2011-Jun	"Big Data technologies describe a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis."	value, new architecture
(Cuzzocrea et al., 2011)	2011-Oct	"enormous amounts of unstructured data produced by high-performance applications falling in a wide and heterogeneous family of application scenarios"	-
(Microsoft, 2012)	2012-Feb	"Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information."	computing power
(Lamont, 2012)	2012-Apr	Volume, variety, velocity	-
(Madden, 2012)	2012-May	"...it means data that's too big, too fast, or too hard for existing tools to process."	-
(Gartner, 2012)	2012-Jun	"Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."	decision making
(Gerhardt et al., 2012)	2012-Jun	Volume, variety, velocity	-
(EMC, 2012)	2012-Jul	"Big data refers to new scale-out architecture that address these needs. [quickly processing more	-

Authors	Date	Definition	New perspective
		varied, more complex, and less structured data] Big data is fundamentally about massively distributed architectures and massively parallel processing, using commodity building blocks to manage and analyze the data."	
(Schneider, 2012)	2012-Sep	Volume, variety, velocity	-
(DeloitteConsulting, 2012)	2012-Oct	"Big data generally refers to datasets so large and complex they create significant challenges for traditional data management and analysis tools in practical timeframes."	practical timeframes
(Frankel, 2012)	2012-Oct	"the volumes of structured and unstructured data produced as a by-product of operating a company"	by-product
(McAfee et al., 2012)	2012-Oct	Volume, variety, velocity	-
(Schroeck et al., 2012)	2012-Oct	"...characterizing three dimensions of big data – "the three Vs:" volume, variety and velocity. And while they cover the key attributes of big data itself, we believe organizations need to consider an important fourth dimension: veracity."	veracity
(Chen et al., 2012)	2012-Dec	big data and big data analytics have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies.	visualization
(Fan and Bifet, 2013)	2012-Dec	Volume, Velocity, Variety, Value, Variability	variability
(Cackett, 2013)	2013-Feb	Volume, velocity, variety, value	-
(Gardner, 2013)	2013-Mar	"Volume, velocity, variety"	-
(O'Leary, 2013)	2013-Mar	"Big data isn't just volume, variety, and velocity, though; it's volume, variety, and velocity at scale."	-
(Provost and Fawcett, 2013)	2013-Mar	For this article, we will simply take big data to mean datasets that are too large for traditional data-processing systems and that therefore require new technologies.	-
(TataConsultancyServices,	2013-Mar	Volume, variety, velocity	-

Authors	Date	Definition	New perspective
2013)			
(Wang et al., 2013)	2013-Mar	" Big Data refers to large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and other digital sources available today and in the future"	distributed data sets
(Demchenko et al., 2013)	2013-May	"...we intend to propose wider definition of Big Data as 5 Vs: Volume, Velocity, Variety and additionally Value and Veracity."	-
(Membrey et al., 2013)	2013-May	"Extensions to the (3V) model that take Value into account are then proposed and discussed. … However recording the data does not bring any value to the company. It only becomes valuable once that data is used or processed. ... High Value Data (HVD) is data that has a known benefit from its storage. ... Low Value Data (LVD) is data that is stored in the anticipation that value will be drawn from it in the future."	High & low value data
(Sagiroglu and Sinanc, 2013)	2013-May	Volume, variety, velocity	-
(Zhang et al., 2013)	2013-Jul	Volume, velocity, variety, value	-
(Bertolucci, 2013)	2013-Aug	Big data is about "building new analytic applications based on new types of data, in order to better serve your customers and drive a better competitive advantage."	competitive advantage
(Ward and Barker, 2013)	2013-Sep	"Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning."	-
(Stonebraker and Robertson, 2013)	2013-Sep	In summary, big data can mean big volume, big velocity, or big variety.	-
(Ferrando-Llopis et al., 2013)	2013-Oct	Volume, velocity, variety, veracity	-
(TheIrishTimes, 2013)	2013-Nov	"A simple definition is that it gives organisations insights into data which they don't already have and does that in a way that helps them improve their operational efficiency and helps them make better decisions."	-
(Vossen, 2014)	2013-Nov	Volume, velocity, variety, veracity	-

Authors	Date	Definition	New perspective
(Balar et al., 2013)	2013-Dec	Volume, variety, velocity	-
(Xin and Ling, 2013)	2013-Dec	Volume	-
(Pospiech and Felden, 2013)	2013-Dec	Volume, variety, velocity	-
(Chen et al., 2014)	2014-Jan	Volume, variety, velocity	-
(Spiess et al., 2014)	2014-Feb	Volume, variety, velocity	-
(Ashraf et al., 2015)	2014-Apr	Volume, velocity, variety, value, veracity	-
(Pandey and Tokekar, 2014)	2014-Apr	Volume, variety, velocity	-
(Blume et al., 2014)	2014-May	Volume	-
(Collins, 2014)	2014-May	Volume, variety, velocity	-
(Demchenko, DeLaat, et al., 2014)	2014-May	"Big Data (Data Intensive) Technologies are targeting to process high-volume, high-velocity, high-variety data (sets/assets) to extract intended data value and ensure high-veracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) new data models (supporting all data states and stages during the whole data lifecycle) and new infrastructure services and tools that allow obtaining (and processing) data from a variety of sources (including sensor networks) and delivering data in a variety of forms to different data and information consumers and devices."	delivery to consumers
(Demchenko, Ngo, et al., 2014)	2014-May	Volume, velocity, variety, value, veracity	-
(Benjamins, 2014)	2014-Jun	Volume, variety, velocity	-
(Hu et al., 2014)	2014-Jun	Volume, velocity, variety, value	-

Authors	Date	Definition	New perspective
(Li et al., 2014)	2014-Jun	" Big Data refers to large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and other digital sources available today and in the future"	-
(Lu et al., 2014)	2014-Jul	Volume, variety, velocity	-
(Meng and Meng, 2014)	2014-Jul	Volume, velocity, variety, value	-
(Richards, 2014)	2014-Aug	Big Data is commonly defined as data that cannot be processed by standard database systems.	-
(De Mauro et al., 2015)	2014-Sep	"Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value."	Information assets
(Lin, 2014)	2014-Nov	Volume, variety, velocity	-
(Akerkar, 2014)	2014-Dec	"Using big volume, big velocity, big variety data asset to extract value (insight and knowledge), further confirm veracity (quality and trustworthiness) of the original data and the acquired information, that demand cost-effective, novel forms of data and information processing for enhanced insight, decision making, and processes control. Additinally, those demands are supported by new data models and new infrastructure services and tools which are able to procure and process data from a variety of sources and deliver data in a variety of forms to several data and information consumers and devices."	-

Authors	Date	Definition	New perspective
(Demchenko, Gruengard, et al., 2014)	2014-Dec	"Big Data (Data Intensive) Technologies are targeting to process high-volume, high-velocity, high-variety data (sets/assets) to extract intended data value and ensure highveracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) new data models (supporting all data states and stages during the whole data lifecycle) and new infrastructure services and tools that allow obtaining (and processing) data from a variety of sources (including sensor networks) and delivering data in a variety of forms to different data and information consumers and devices."	-
(Hashem et al., 2015)	2015-Jan	Volume, velocity, variety, value	-
(Tiefenbacher and Olbrich, 2015)	2015-Jan	Volume, variety, velocity (in), visibility, veracity, virtue (= value), velocity (out)	velocity in, velocity out
(Baro et al., 2015)	2015-Feb	"Volume: Log(n * p) ≥ 7" (n=statistical individuals, p=nbr of variables) Properties: "Great variety, High velocity, Challenge on veracity, Challenge on all aspects of the workflow, Challenge on computational methods, Challenge on extracting meaningful information, Challenge on sharing data, Challenge on finding human experts"	workflow
(Jin et al., 2015)	2015-Feb	Volume, velocity, variety, value, veracity	-
(Wyber et al., 2015)	2015-Mar	Volume, velocity, variety, veracity	-
(Gandomi and Haider, 2015)	2015-Apr	Volume, variety, velocity	-
(Emani et al., 2015)	2015-May	Volume, velocity, variety, veracity	-