Defining Big Data

Small data are slow and sampled. Big Data are quick and n=all.

Kitchin & McArdle (2016)

This chapter searches for defining properties of big data, focusing on characteristics with possible implications for cartographic practice. Review of related works outlines the main attitudes towards grasping the concept.

Ontological characteristics ("big" standing for "fast" and "exhaustive")

Despite the lively interest triggered by the subject, the explanation of the term *big data*[1] remains hazy and there is no widely accepted definition to the date. Perhaps the most systematic effort in this matter by Kitchin (2014) (refined in Kitchin & McArdle (2016)) summarizes the key properties attributed to big data. Kitchin critically assesses these properties and goes on to assign them relative importance in distinguishing big from "small" data. He also takes care to separate the concept in itself from accompanying social phenomena, hence he speaks of *ontological* characteristics.

This provides a useful starting point for our thinking of big data from the cartographic standpoint, so let us list the ontological characteristics including some of the Kitchin's comments:

- **Volume** – can be measured in storage requirements (terabytes or petabytes) or in number of records
- **Velocity** – data generation happens in real-time either constantly (e.g. CCTV) or sporadically (e.g. web search), we can distinguish the frequency of generation from the frequency of data *handling*, *recording*, and *publishing*, all three can be delayed from the time of generation
- **Variety** – heterogeneous in nature, though this property is rather weak as various levels of organization are allowed (*structured*, *semi-structured* or *unstructured*)

---

1 Big Data vs big data, data in plural or singular?

- **Exhaustivity** – an entire system is captured (*n=all*), rather than being sampled
- **Resolution and indexicality** – fine-grained (in resolution) rather than being aggregated and uniquely indexical (in identification) which enables linking with other datasets
- **Relationality** – containing common fields that enable the conjoining of different datasets
- **Extensionality and scalability** – flexibility of data generation, possibility to add or change new fields easily, possibility to rapidly expand in size

The above parameters, their relative importance, as well as their cartographic implications, will be discussed more in depth later, but straight from the start it is important to mention two open questions that for many people make attempts to define big data vague at best, sometimes to the point of questioning the existence of the phenomenon itself.

First, there are no quantitative thresholds that would define exactly how large the "big" volume is, how fast the "big" velocity is, and so on. Some properties would even be hard to describe in quantitative terms (for example extensionality). Other properties sound too general or vague to act as a sound defining parameter (scalability). What is more, one could extend the properties ad absurdum, for example *variety* could refer to differences in structure, origin, quality, or any other property of a dataset. Such depth of parameters and sub-parameters does not add to the overall comparability of datasets, especially when we consider that data-generation procedures may be unique to certain domains and not found in others. Finally, many datasets lack metadata detailed enough to allow to judge all mentioned properties. It is possible that these issues will clear out with time, but parameter thresholds may as well remain blurry and ever in flux.

The second problem is that even if we had a clearly defined set of criteria, in practice we could hardly find a dataset that would fit all of them. Therefore not all properties are deemed mandatory, which in turn leads to confusion and labeling almost anything as big data. To articulate the gist of the term, more work is needed on the relations of the parameters, some might be merged (resolution is a consequence of exhaustivity, indexicality enables relationality) or discarded (extensionality and scalability seem to describe the infrastructure rather than data).

Aware of these problems, Kitchin & McArdle (2016) argues that *velocity* and *exhaustivity* are qualities that set big data apart and distinguish them from "small" data. We can add that these two characteristics also present the most interesting challenges to cartographic presentation of such data. So even though we will continue to use the established term in the following text, the little too simplistic adjective "big" will be meant as a proxy for **generated continuously in real time and containing an unreduced set of elements**.

Other ways of understanding big data

In this section we briefly review the writing of authors striving to define big data. The term itself was fist used in context of dealing with massive datasets in mid-1990s by John Mashey (F. X. Diebold et al. (2012)), but the heaviest circulation of the term in scientific and popular media takes place only in recent years. From the breadth of works, several tendencies can be identified, providing more or less illuminating interpretations of the subject (for an alternative summary of definitions see Gandomi & Haider (2015), for bibliometric analysis of related scientific literature see Nobre & Tavares (2017)).

Vs and keywords

Kitchin's taxonomy mentioned in the previous section is based on a review of older definitions, starting with the often-cited three Vs (standing for *volume*, *velocity*, and *variety*) by Laney (2001). The notion of *exhaustivity* was added by Mayer-Schönberger & Cukier (2013), concepts of *resolution* and *indexicality* came from Dodge & Kitchin (2005), Boyd & Crawford (2012) adds *relationality*, and the qualities of *extensionality* and *scalability* were taken from Marz & Warren (2012).

Other properties attributed to big data include *veracity* (data can be messy, noisy and contain uncertainty and error) and *value* (many insights can be extracted, data can be repurposed), both brought forward by Marr (2014). Moreover, *variability* (the meaning obtainable from data is shifting in relation to the context in which they are generated) was identified by David Hopkins in relation to text analysis (Brunelli (2011)).

Laney's 3 Vs were brought into commercial management-speak and became a slogan further powering the hype of big data. Nevertheless, it inspired a number of other authors to extend it quite creatively. For example Uprichard (2013) lists other v-words to be considered, both in positive (*versatility*, *virtuosity*, *vibrancy*...) and negative (*valueless*, *vampire-like*, *violating*...) light. Marr (2014) describes five Vs of big data, Van Rijmenam (2013) sees seven Vs, Boellstorff & Maurer (2015) propose three Rs and Lupton (2015) even uses thirteen p-words to describe the subject. As Kitchin & McArdle (2016) notes, these additional v-words and new p-words are often descriptive of a broad set of issues associated with big data, rather than characterising the ontological traits of data themselves.

A challenge for technical infrastructure

Several authors understand big data mainly as a management issue, which is probably due to the fact that handling large datasets is challenging. Hence, the computational difficulties of storing and processing a dataset on a single machine often act as a defining measure. One example for all is Storm (2012) quoting Hillary Mason: "Big Data usually refers to a dataset that is too big to fit into your available memory, or too big to store on your own hard drive, or too big to fit into an Excel spreadsheet."

The problem with such definitions is determining exactly what size is "too big to fit". The computational power of hardware accessible for personal use is constantly increasing (by the Moore's law[2]), not to mention the technical infrastructure accessible to large enterprises and governmental organizations (datacenter construction is steadily growing and is expected to almost double the current capacity in 2021 ("Data center storage capacity worldwide from 2016 to 2021, by segment (in exabytes)," 2018), Networking (2018)).

At the same time, new technologies emerge to address the issue – virtualization of storage, networking, and memory make it possible to rent computational infrastructure from "cloud" providers, or to delegate workloads previously carried out by the operating system to remote platforms.[3] Other innovations take place in data processing algorithms,

---

2 Moore's law – explanation, critique (this idea must die)

3 *Cloud computing* enables companies to consume a compute resource, such as a virtual machine, storage or an application, as a utility rather than having to build and maintain computing infrastructures in house (Rouse (2018)). The cloud models include providing

analytic engines, and in database design (a whole range of No-SQL databases as well as enablement of distributed processing in traditional databases).[4] Some attempts to summarize technical solutions for big data can be found in Pääkkönen & Pakkala (2015), or Jin, Wah, Cheng, & Wang (2015).

As we can see, the "too big to fit" definitions are highly dependent on the resources currently available, plus we need to take into account future improvements that are hard to predict. That being said, understanding the subject as *data that prevent local offline processing on common desktop in reasonable time* is a useful shorthand for judging big from "small" data. The threshold between local (offline) and remote (cloud-dependent) processing exists even though it is a blurry and a dynamic one. As the latter may be more and more accessible in the future, it can be best advised to consider the scalability of any data-processing workflows early on. In other words, any workflow designed as a potential big data process will likely have an advantage, as design limitations may prove to be overcome harder than the technical ones.

Showing example sources and quantities

A very common description of big data goes along the lines of "I will give you some numbers and you will get what I mean". Such writing may not provide an exact understanding of the concept, but can put us into context about the scales we are moving at. Doubtlessly the mass of retained data is growing, as McNulty (2014) put it, 90% of all data ever created was generated in the past 2 years (in 2014). In a notable attempt to estimate the World's overall data generation between 1986 and 2007, Hilbert & López (2011) claim that more then 300 exabytes[5] of stored data existed in 2007 (for the methodology of reckoning see Hilbert & López (2012)). The key insight is the growing domination of digital technologies accounting for the majority of the annual growth after year 2000. More recent accounts report on machines potentially capable of reporting brontobytes[6] of data (Bort (2014)).
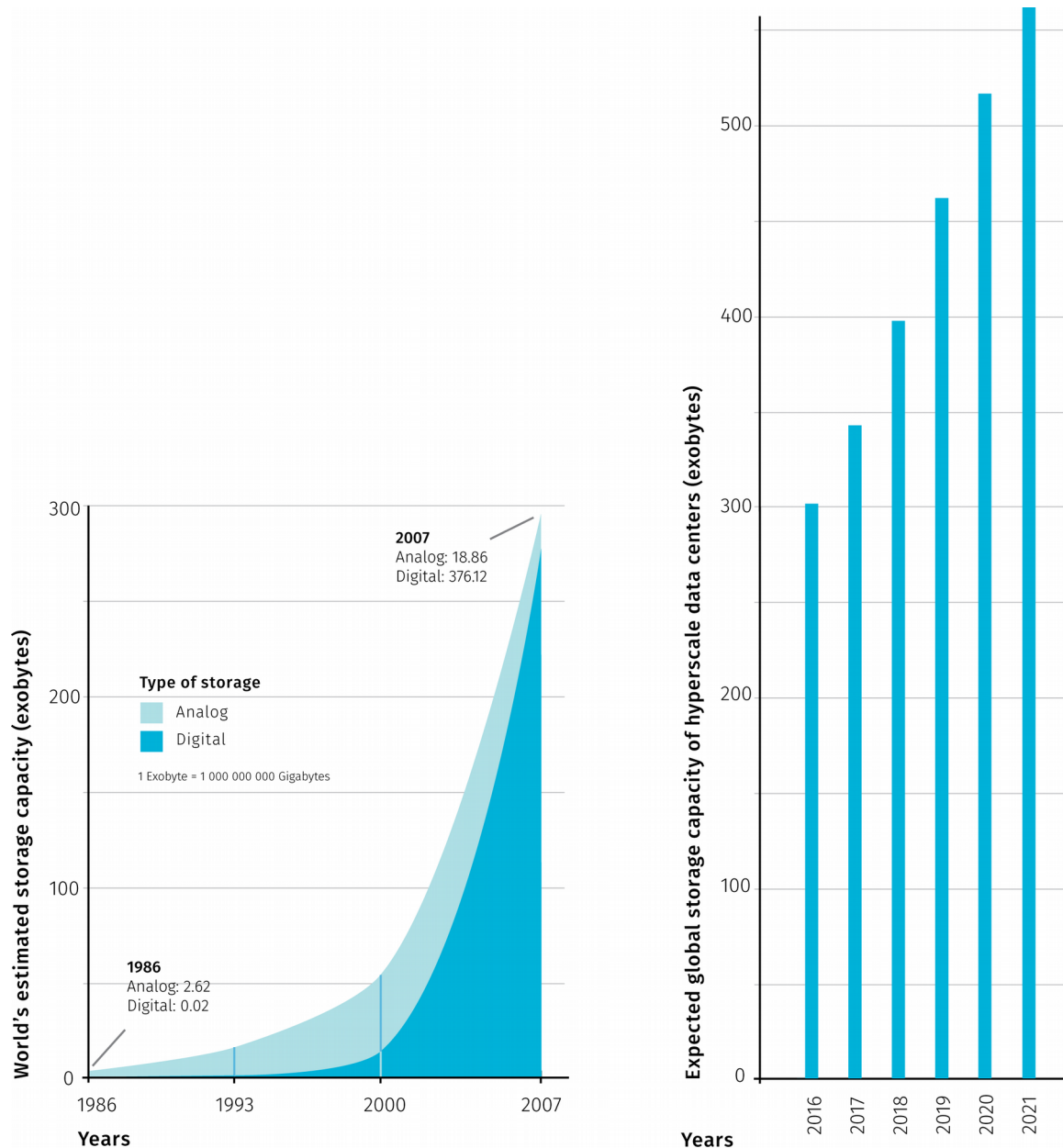
---

infrastructure, platform or application as a service; main vendors of public cloud solutions are Amazon Web Services, Google Cloud Platform or Microsoft Azure.

4 Processing and analytical frameworks designed for big data inclued Apache Hadoop, Apache Spark, or Apache Flink. No-SQL databases use a collumn, graph, document, key-value, or multi-model solution as an alternative to tradional reltional database design.

5 1 exabyte = 1 000 000 000 gigabytes

Increasing the storage capacity itself does not speak of any qualitative change in what is stored, therefore some archives could indeed be described as big piles of small data. Under certain circumstances, new quality can arise from increased quantity, for example as Norvig (2011) points out, an array of static images projected at a sufficient frame rate creates an illusion of movement also known as film. Multiplication of an old medium creates a new one. The remaining question is under what conditions this change of essence arises, and if such thing occurs or will occur in case of big data.

Type of storage
- Analog
- Digital

1 Exobyte = 1 000 000 000 Gigabytes

2007
Analog: 18.86
Digital: 376.12

1986
Analog: 2.62
Digital: 0.02

World's estimated storage capacity (exobytes)

Years

Expected global storage capacity of hyperscale data centers (exobytes)

Years

---

6 1 brontobyte = 1 000 000 000 exabytes

Fig.1 *Comparison of the World's estimated data storage capacity between years 1968 and 2007 (modified after Hilbert & López (2011)) and the expected storage capacity of large scale data centers in the period from 2016 to 2021 (modified after Networking (2018))*

Rather than putting up to a gargantuan task of counting the mass of all existing data items, authors use the available statistics related to operations of large companies (Kambatla, Kollias, Kumar, & Grama (2014), McNulty (2014), Marr (2014) and others). For example, Facebook was said to process 10 billion messages, 4.5 billion button clicks and 350 million picture uploads each day (Marr (2014)). It goes without saying these numbers are outdated and certainly outgrown today. Other companies prominently mentioned in context of big data are Google, Wallmart, or Amazon. This connection is justified, as these companies have put user (or customer) data analytics to the core of their businesses, thus supporting the progress in the field. Social media, web search and browsing data, on-line or off-line shopping patterns, but also mobile devices, sensors and large scientific projects are mostly named as generators of big data.

Data source is another potential classification property. The United Nations Economic Commission for Europe proposed a taxonomy that recognizes three main sources (UNECE (2013)):

- *Social Networks (human-sourced information)* – this information is the record of human experiences
- *Traditional Business systems (process-mediated data)* – these processes record and monitor business events of interest
- *IoT (machine-generated data)* – information is derived from sensors and machines used to measure and record the events and situations in the physical world

Data sources labeled as big differ from traditional sources such as surveys and official administrative statistics, Florescu et al. (2014) and Kitchin (2015) closely examine those differences as well as the potential for big data to extend official statistics. Interesting point is that volume is not actually distinctive as statistic offices tend to store large amounts as well. Classical data sources have statistical products and by-products specified beforehand, big data tend to be reused beyond the original intent, on the other hand, big data sources tend to be volatile and their representativeness is harder to assess.

Metaphoric accounts

Metaphors rely on a notion of analogy between two dissimilar things, but can also become independent verbal objects, aesthetically appealing but not overly revealing. Despite that, we should not ignore metaphoric accounts, as they contribute to the mythology surrounding big data that reflects what many people expect.

Puschmann & Burgess (2014) identified two prevailing ways of imagining the subject: big data seen as a natural force to be controlled and as a resource to be consumed. The utilitarian mindset comparing digital world to excavation of valuable minerals in far from new (think of "data mining" or more recently "cryptocurrency mining") bur it is tempting pursue this analogy further. For example, how to estimate the ratio of valuable information to "debris", and shouldn't such estimation be done before any data "mining" endeavour? The value of real-world analogies may be in provoking some common-sense reasoning often missing in proclamations of digital visionaries.

As Lupton (2013) notes, by far the most commonly employed rhetorical descriptions of big data are those related to water or liquidity, suggesting both positive and negative connotations. For example Manyika et al. (2013) argues for unlocking data sources to become "liquid" in a sense of open and free-flowing, at the same time keeping privacy concerns in mind – what is liquid is also susceptible to unwanted leaks.

Of course, big data descriptions are not limited to verbal form, visual means can be much more expressive and informative – not a surprising claim to be found in a thesis on visual analytics. We will discuss cartographic tools later, here we can mention artistic renderings that employ more free-form visual analogies. We should distinguish pursuits like *information visualisation* that are close to graphic design (for good overview see klanten2010data or Lima (2011)) from artistic projects that use data as a raw material and don't aim to convey information or comfort to general user's cognitive expectations (like some projects at ("Creative applications network," 2018)).

From the cartographer's standpoint, aspects of visual art can be inspiring (graphic quality, employment of computation and rendering software, creative uses of interaction and animation), though artistic means are often too different to be transposed. Without referring back to the source

phenomenon, data-driven art becomes unrecognizable from the so called *generative art* that uses random numbers rather than any existing information (though generative approaches and data-mocking may prove some use for web cartography).

Holistic and process-oriented accounts

Multifaceted phenomena tend to provoke descriptions that contain words like "system" or "process". Somewhere behind that lies an idea that a whole more than a sum of its parts[7]. Experts from different fields notice aspects of phenomena that are close to their priorities, interdisciplinary definitions combine such views. Naturally, listing holistic accounts will include topics already mentioned, so pardon some repetition in this section.

Murthy, Bharadwaj, Subrahmanyam, Roy, & Rajan (2014) prepared a taxonomy of big data comprised of:

- *data* – with various levels of temporal latency and structure
- *compute infrastructure* – batch or streaming
- *storage infrastructure* – sql, nosql or newsql
- *analysis* – supervised, semisupervised, unsupervised or re-enforcement machine learning
- *visualisation* – maps, abstract, interactive, real-time
- *privacy and security* – data privacy, management, security

On the other hand, Boyd & Crawford (2012) define big data as a cultural, technological, and scholarly phenomenon that rests on the interplay of:

- *technology* – maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets
- *analysis* – drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims
- *mythology* – the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy

As the two taxonomies above illustrate, there are many ways to slice a cake. The fate of overreaching definitions is being too intricate to explain the

---

7 This is true for systems that express *synergy* or *emergent behavior*. Over the years, *systems theory* has been used to describe many natural and man-made phenomena in many fields, including geography and cartography.

phenomena crisply, yet never complete as there is always a point of view that hasn't been included. So here we arrive at a trade-off between preciseness of a definition and its practicality.

One way out from this is simply rejecting the view of big data as a singular phenomenon. Big data is a non-specific covering term that could mean different things to different people. As Helles & Jensen (2013) observes, "[d]ata are made in a process involving multiple social agents — communicators, service providers, communication researchers, commercial stakeholders, government authorities, international regulators, and more. Data are made for a variety of scholarly and applied purposes [...]. And data are processed and employed in a whole range of everyday and institutional contexts."

Spatial properties of big data

Apart from the general definitions mentioned above, there have also been field-specific efforts to contextualize big data. The fields include governance (Crampton (2015)), journalism (Lewis & Westlund (2015)), ecology (Shin & Choi (2015)), social sciences (ovadia2013role), or business administration (wamba2015big), learning analytics (Wilson, Thompson, Watson, Drew, & Doyle (2017)), education (kabakchieva2015big) and doubtlessly many others.

how to contextualize bd within geography – maybe to the next section on science: Graham & Shelton (2013) Kitchin (2013)

Geospatial big data as a subset of big data.

Lee & Kang (2015) Li et al. (2016)

A potential to geocode big data to become available to geospatial analysis.

Cartography (here represented by ICA) is no exception. ... Manifestos, etc...

Assessing impacts and opportunities rather than seeking definitions

Often times, big data are described indirectly by the impacts, real or imagined, it has on the society. For some authors, the debate on definition of big data may be dismissed as unproductive. The popularity of the term

itself may diminish like many other new technologies that become part of the infamous hype cycle.[8] Many ideas in the IT industry exist under changing or concurrent names, and big data has indeed a lot in common with terms such as *data mining*, *business intelligence* or *visual analytics* to name just a few. But we should not forget that even though the technological industry is largely fashion-driven, its societal impacts are real, though sometimes unevenly distributed.

The definition of big data is elusive perhaps also because the majority of involved actors,as they are positioned in the business world, is more focused on building productive big data ventures, without much conceptual attention to the subject in itself. Then of course, the underlying technologies become a subject of marketing which often uses inflated overstatements based on expectations rather than reality. So far there is no settled consensus around Big Data definition in the academia either, but as Kitchin & McArdle (2016) predicts, the "genus" of big data will probably be further delineated and its various "species" identified. The question is if the then such an umbrella term will be necessary. Anyways, the lack of common ground in understanding what big data is (illustrated by this chapter) may be a good predictor of the term's future relevance. Problems with definition is exactly what leads Davenport (2014) to "predict a relatively short life span for this unfortunate term."

To mayer2013big big data stands for "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value". Here, more than an exact definition, the important lies in the real-life impacts that are likely to stay even when the big data hype is over. Even if we dismiss big data as a buzz-word, the fact is that more digital information gets created and it can be linked more easily, which has many implications on the way we live. Together with that there are changing attitudes to putting data to work. Before we move to describing the challenges and opportunities big data pose to cartography, we dedicate the following chapter to some societal and scientific impacts of the phenomenon, as they can provide a motivation for cartography to take part in addressing those issues.

_____

8 Hype cycles describe how expectations from emerging technologies evolve with time. Stages in the cycle are: *innovation trigger*, *peak of inflated expectations*, *trough of disillusionment*, *slope of enlightenment*, and *plateau of productivity*. The expected duration of cycle differs per technology, and some technologies may never reach productivity. Hype cycles are an invention of consultancy Gartner, that issues regular reports Gartner (2018)

Sources

Boellstorff, T., & Maurer, W. (2015). Introduction. In *Data, now bigger and better!* (pp. 1–6). Prickly Paradigm Press.

Bort, J. (2014). There's a new word being used in the computer industry: 'Brontobytes'. *Available online at http://www.businessinsider.com/new-big-data-word-brontobytes-2014-6 (last accessed May 30, 2018)*.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662–679. Taylor & Francis.

Brunelli, M. (2011). Will your organization benefit from 'big data' processing technology? *Available online at searchdatamanagement. techtarget. com/news/2240036228/Will-your-organization-benefit-from-big-data-processing-technology (last accessed December 29, 2016)*.

Crampton, J. W. (2015). Collect it all: National security, big data and governance. *GeoJournal*, *80*(4), 519–531. Springer.

Creative applications network. (2018). *Available online at http://www.creativeapplications.net/ (last accessed May 30, 2018)*.

Data center storage capacity worldwide from 2016 to 2021, by segment (in exabytes). (2018). *Available online at https://www.statista.com/statistics/638593/worldwide-data-center-storage-capacity-cloud-vs-traditional/ (last accessed May 30, 2018)*.

Davenport, T. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Harvard Business Review Press.

Diebold, F. X., Cheng, X., Diebold, S., Foster, D., Halperin, M., Lohr, S., Mashey, J., et al. (2012). A personal perspective on the origin (s) and development of "big data": The phenomenon, the term, and the discipline∗. Citeseer.

Dodge, M., & Kitchin, R. (2005). Codes of life: Identification codes and the machine-readable world. *Environment and Planning D: Society and Space*, *23*(6), 851–881. SAGE Publications.

Florescu, D., Karlberg, M., Reis, F., Del Castillo, P. R., Skaliotis, M., & Wirthmann, A. (2014). Will 'big data'transform official statistics. In *Q2014– European conference on quality in statistics*.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. Elsevier.

Gartner. (2018). Gartner special reports. *Available online at https://www.gartner.com/en/products/special-reports (last accessed August 26, 2018)*.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, *3*(3), 255–261. SAGE Publications Sage UK: London, England.

Helles, R., & Jensen, K. (2013). Making data—big data and beyond: Introduction to the special issue. *First Monday*, *18*(10).

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, *332*(6025), 60–65. American Association for the Advancement of Science.

Hilbert, M., & López, P. (2012). How to measure the world's technological capacity to communicate, store, and compute information part i: Results and scope. *International Journal of Communication (19328036)*, *6*.

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

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561–2573. Elsevier.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, *3*(3), 262–267. Sage Publications Sage UK: London, England.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, *31*(3), 471–481. IOS Press.

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 2053951716631130. SAGE Publications.

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

Lee, J.-G., & Kang, M. (2015). Geospatial big data: Challenges and opportunities. *Big Data Research*, *2*(2), 74–81. Elsevier.

Lewis, S. C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, *3*(3), 447–466. Taylor & Francis.

Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., Pettit, C., et al. (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 119–133. Elsevier.

Lima, M. (2011). Visual complexity. mapping patterns of information. Princeton: Princeton Architectural Press.

Lupton, D. (2013). Swimming or drowning in the data ocean? Thoughts on the metaphors of big data. *Available online at https://simplysociology. wordpress. com/2012/10/29/swimming-or-drowning-in-the-data-ocean-thoughts-on-the-metaphors-of-big-data/ (last accessed December 29, 2016)*.

Lupton, D. (2015). The thirteen ps of big data. *Available online at https://simplysociology. wordpress. com/2015/05/11/the-thirteen-ps-of-big-data/ (last accessed December 29, 2016)*.

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. *McKinsey Global Institute*, 21.

Marr, B. (2014). Big data: The 5 vs everyone must know. *LinkedIn. Available online at www. linkedin. com/pulse/20140306073407–64875646-bigdata-the-5-vs-everyone-must-know (last accessed December 29, 2016)*.

Marz, N., & Warren, J. (2012). *Big data: Principles and best practices of scalable realtime data systems*. MEAP Edition Manning Publications Co.

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

McNulty, E. (2014). Understanding big data: The seven v's. *Available online at dataconomy. com/2014/05/seven-vs-big-data/(last accessed December 29, 2016)*.

Murthy, P., Bharadwaj, A., Subrahmanyam, P., Roy, A., & Rajan, S. (2014). Big data taxonomy. *Cloud Security Alliance (CSA)*, *Tech. Rep*.

Networking, C. V. (2018). Cisco global cloud index: Forecast and methodology, 2016-2021. white paper. *Cisco Public*, *San Jose*.

Nobre, G. C., & Tavares, E. (2017). Scientific literature analysis on big data and internet of things applications on circular economy: A bibliometric study. *Scientometrics*, *111*(1), 463–492. Springer.

Norvig, P. (2011). The unreasonable effectiveness of data - ubc distinguished lecture series. *Available online at https://www.youtube.com/watch?v=yvDCzhbjYWs (last accessed May 30, 2018)*.

Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big Data Research*, *2*(4), 166–186. Elsevier.

Puschmann, C., & Burgess, J. (2014). Big data, big questions| metaphors of big data. *International Journal of Communication*, *8*, 20.

Rouse, M. (2018). Cloud computing. *Available online at https://searchcloudcomputing.techtarget.com/definition/cloud-computing (last accessed May 30, 2018)*.

Shin, D.-H., & Choi, M. J. (2015). Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, *32*(2), 311–320. Elsevier.

Storm, D. (2012). Big data makes things better. *Available online at insights. dice. com/2012/08/03/big-data-makes-things-better/ (last accessed December 29, 2016)*.

UNECE. (2013). UNECE - united nations economic commission for europe. *Available online at https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data (last accessed August 26, 2018)*.

Uprichard, E. (2013). Focus: Big data, little questions? *Discover Society*, (1). Social Research Publications.

Van Rijmenam, M. (2013). Why the 3v's are not sufficient to describe big data. *Available online at http://www. bigdata-startups. com/3vs-sufficient-describe-big-data (last accessed December 29, 2016)*.

Wilson, A., Thompson, T. L., Watson, C., Drew, V., & Doyle, S. (2017). Big data and learning analytics: Singular or plural? *First Monday*, *22*(4).