The next wave: "big data"? (*)

Giuditta DE PRATO
JRC EC IPTS, Sevilla, Spain

Jean Paul SIMON
JPS Public Policy Consulting, Sevilla, Spain

Abstract: The paper introduces an overview of the issue of "big data". It provides some "guestimates" of its size (volume, market, features), and investigates the real meaning of the phenomenon based on case studies from big data pioneers and involved industries. It takes a look at the main drivers, and at some likely outputs. The paper identifies some of the tensions and issues brougth by this new growth industry and reviews the policies needed to face the challenges ahead.

Key words: analytics 3.0, big data value chain, business intelligence, data management data in the cloud, digital dragons, internet of things, machine-to-machine communication, metrics, value added content providers.

This article aims at marshalling facts about a notion that have been spreading quickly without being, most of the times, properly defined thereby remaining all-encompassing. A fully comprehensive analysis of the phenomenon is clearly out of the scope of such an article, however this article tries to put the phenomenon into perspective, stressing the main challenges ahead: the economic, business and policy challenges. It is meant to provide an introduction based on a review of the main sources available. The paper is based on desk research, a review of literature, review of the technical journals, annual reports, company websites and other internet sources. It also builds on the authors' own researches and experience.

The first section gauges the size of the data involved; it offers glimpses of some of the assessment of the (potential) market linked to "Big data" [1]. The section offers a tentative definition of what the somewhat fuzzy notion

(*) The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of their institutions.

1 From here on big data without quotation marks.

covers. The second section questions the substance of the phenomenon trying to better underline its real present scope, to investigate beyond the present hype. The third section then aims at explaining what the phenomenon really means, fleshing out it real content through some field applications stemming from various industries. The last section deals with the threats and tensions it may bring along but also with the opportunities in several areas. It reviews some of the initiatives taken by governments. The paper concludes delineating some potential policy interventions.

■ How "big" is big? From MegaBytes to Zettabytes [2]

The best "guestimates" [3] of the total amount of data in the world suggest that from 1987 to 2007 the total amount of analogue and digital data in the world grew from 3 billion gigabytes to 300 billion gigabytes, an increase by a hundred in two decades (HAIRE & MAYER-SCHÖNBERGER, 2014:5). The so-called data explosion is driven by the combination of exponentially expanding amount of data available (up to 7 "zettabytes" predicted for 2015), and the rapidly improving ability to process and analyse the data (Boston Consulting Group, 2012: 7). It is enabled by the deployment of the relevant infrastructure (networks and devices)

The data originate from various and heterogeneous different sources like people, machines or sensors. As emphasized by the UN Global Pulse, big data is both the information that is passively generated as by-products of people's everyday use of technologies and the information people willingly communicate about themselves on the web [4]. "Data fusion", brings together disparate sources of data, "digital data" and "analogue data" (emanating from the physical world, but increasingly converted into digital format).

One of the simplest indicators of the growth is the dramatic increase of the mobile traffic data that Cisco is monitoring, with its Cisco Visual

---

[2] MegaByte, 106 bytes, GigaByte, 109 bytes, PetaByte: 1015 bytes, ExaByte, 1018 bytes, Zettabytes 1021bytes.

[3] One should be very careful about the data on "big data", the figures collected and provided in this section are just meant to give some indications about trends. See the 2nd section for the lack of definition of the core terminology. As noted by the eSkill UK /SAS 2013 report: "reported adoption rates vary significantly, and in most cases observed are subject to significant caveats not always readily highlighted within the associated study documents" (.e-skills UK/ SAS, 2013).

[4] http://www.unglobalpulse.org/about/faqs

Networking Index, signalling at the same time the dominance of video (nearly 79% of the total traffic predicted for 2018, Cisco VNI 2014: 14), and the leading role of consumers.

According to the EMC-IDC annual study (EMC Digital Universe study, 2014) the data created and copied annually will reach 44 zettabytes, or 44 trillion gigabytes. What they call the "digital universe" (see Figure 1) is growing 40% a year into the next decade, not only the increasing number of people and enterprises doing everything online, but also all the "things" (i.e. smart devices) are included.

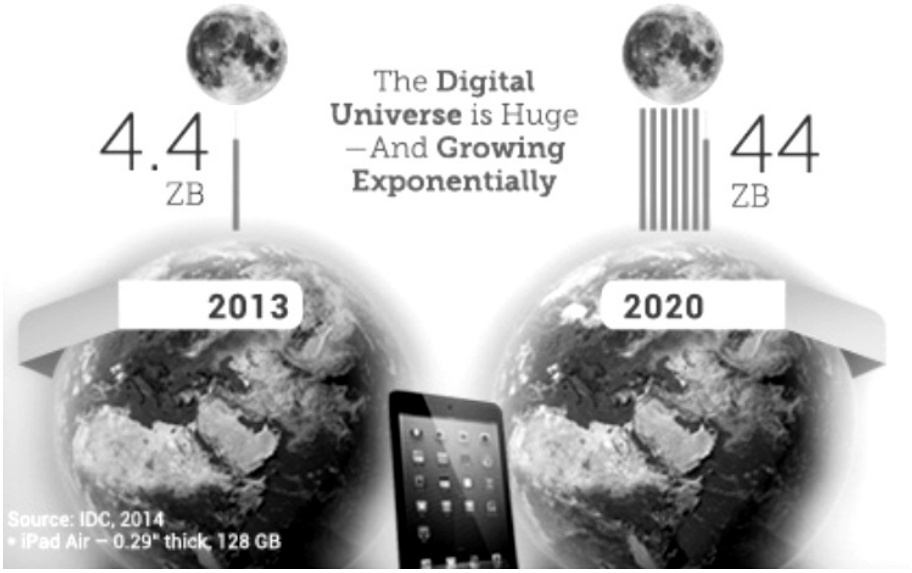
Indeed, the combination of the devices and the relevant networks paves the way for a faster growth of the "internet of things" (IoT) that will generate huge amounts of data. The pervasive integration of semiconductors, mobile communication, and "big data"/analytics is held as a way of propelling the internet of things into the wider economy. The growth of machine-to-machine (M2M) communication is contributing to this expansion of data produced. According to the Ericsson Mobility Report (2014:28), M2M communication is taking off, driven by declining costs, improved coverage, more capable radio technologies, regulatory mandates and a growing range of successful applications and business models. The 2014 Ericson report estimates that, at the end of 2013, there were around 200 million cellular M2M devices in active use; the number is expected to grow 3-4 times by 2019. SAP (2014) predicts 2.1 billion connected devices worldwide by 2021 (SAP, 2014: 6).

The global market for "big data" was estimated at 6.3 billion US $ in 2012 and is expected to reach 48.3 billion US $ by 2018 (a CAGR of 40.5% from 2012 to 2018) (Transparent Market Research, 2013). Another specialised consultancy, IDC predicts that the market for big data will already reach 16.1 billion US $ in 2014 [5], growing 6 times faster than the overall IT market (quoted by Forbes, 2013). According to IDC, the "big data" market will grow to 32.4 billion US $ by 2020 (with an astonishing 212 billion devices connected up to the internet of things) (quoted by 2014). IDC also predicted that by sub-segments 2020, the entire ICT industry will spend 5 trillion US $, over 1.3 trillion US $ more than it does today, from which, "40% of the industry's revenue, and 98% of its growth will be driven by 3rd Platform technologies [6] that today represent just 22% of ICT spending." (IDC, 2014).

[5] IDC includes in this figure infrastructure (servers, storage, etc., the largest and fastest growing segment at 45% of the market), services (29%) and software (24%).

[6] They define the "3rd Platform" as built on the mobile network and apps market, cloud services, social media technologies, and big data analytics.

Figure 1 - The digital universe



Source: IDC, 2014

Big data can be traced back to a Meta Group [7] (2001) report identifying a three dimensional data growth (3Vs) that is often quoted by other reports: volume (amount of data: petabytes or above), velocity (speed of data in and out needed for real-time collection/analysis of data), and variety (range of data types, formats and services, collected from a variety of collection mechanisms). Other sources are now adding veracity (the believability of the data itself) (HAIRE & MAYER-SCHÖNBERGER, 2014: 6), visualization (TASCON, 2013), variability (temporal data peaks) and complexity (issues relating to linking/cleaning/editing data from different sources) (e-skills UK/ SAS, 2013:7).

BAIN (2013a) describes big data, as "the mining and processing of petabytes worth of information to gain insights into customer behaviour, supply chain efficiency and many other aspects of business performance" (BAIN, 2014a: 1). Mc Kinsey (2011: 1) states that big data in many sectors will range from a few dozen terabytes to multiple petabytes (thousands of terabytes). The order of magnitude varies according to the source.

[7] Formerly known as Gartner Group before 2001.

Box 1 - The big data value chain

The big data value chain can be broken down in the following fashion:

- Data Acquisition: Structured data, Unstructured data, Event processing, Sensor networks, Protocols, Real-time, Data streams, Multimodality.

- Data Analysis: Stream mining, Semantic analysis, Machine learning, Information extraction, Linked Data, Data discovery, Whole world semantics, Ecosystems, Community data analysis, Cross-sectorial data analysis.

- Data Curation: Data Quality, Trust / Provenance, Annotation, Data validation, Human-Data Interaction, Top-down/Bottom-up, Community / Crowd, Human Computation, Curation at scale, Incentivisation, Automation, Interoperability.

- Data Storage: In-Memory DBs, NoSQL DBs, NewSQL DBs, Cloud storage, Query Interfaces, Scalability and Performance, Data Models, Consistency, Availability, Partition-tolerance, Security and Privacy, Standardization.

- Data Usage: Decision support, Prediction, In-use analytics, Simulation, Exploration, Visualisation, Modelling, Control, Domain-specific usage.

Source: BIG, http://big-project.eu/sites/default/files/BIG_Introduction.pdf

The basic components of "big data" include software, hardware, and storage, with software and service being the larger share (IDC 2012: 3). New tools to deal with the data (extract, load, and transform) are emerging. Big data is grounded in technologies like Apache Hadoop and NoSql (Not only SQL) (AKAMAI, 2014). The first one is an architecture platform, an open-source software framework that supports data-intensive distributed applications running on large clusters of commodity hardware. The second is a selective data based system for data to be retrieved easily, offering a quick and easy access. NoSQL databases are next generation databases, often "non-relational", distributed and open-source as well as being horizontally scalable.

■ Just another hype?

In 2011, a widely read McKinsey report on big data seems to have triggered part of what can be looked upon as some kind of hype. Other consultancies rode the same wave, pumping up figures, urging the data-naïve to wise-up. However, the McKinsey report was cautious enough, providing a first tentative overview, dealing with selected case studies to illustrate the scope of activities and sectors implied.

To put it more bluntly like HARFORD (2014):

> "As with so many buzzwords, 'big data' is a vague term, often thrown around by people with something to sell".

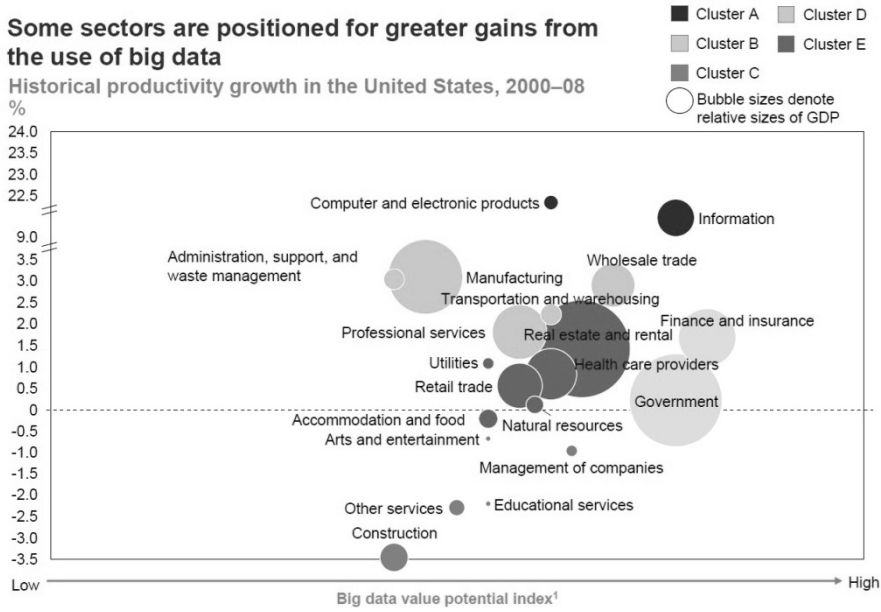
Indeed, there is no agreed definition of big data, no singular, internationally recognised definition of what constitutes 'big data', despite attempts and work being done at the ITU (ITU, 2013 [8]). The terminology employed for the description is not an operational one and, as such, it makes it difficult to come up with clear views about what could constitute a distinct sector, occupation, process, etc. Hence the variations that are to be found among the reports. One way to flesh out the real meaning of big data is to document the way it is being dealt with through case studies, as we will see in the next section.

The McKinsey report already underlined that all sectors were far from being equal, that the propensity and likeliness of adoption vary. Focusing on the US, the report identifies (Figure 2, McKinsey, 2011: 7) four clusters. The "usual suspects" (electronic products and information sectors) are sectors that are set to gain substantially through access to huge pools of data (e.g., internet companies collect vast amounts of online behaviour data). On the opposite, several sectors like construction, educational services, and arts and entertainment, "have posted negative productivity growth", the reports adds that it "probably indicates that these sectors face strong systemic barriers to increasing productivity". In other words, this will likely yield some "laggards" which may clearly trigger some difficult policy issues in order to be in a position to reap the benefits for society at large (4th section). IDC commented that the benefits of big data are not always clear today (FORBES, 2013). Besides, of the useful data, IDC estimates that in 2013 perhaps 5% was especially valuable, or "target rich."(EMC IDC, 2014); adding in a more optimistic mode that:

> "Percentage should more than double by 2020 as enterprises take advantage of new big data and analytics technologies and new data sources, and apply them to new parts of the organization".

[8] For further information on the work on big data carried out by the ITU Telecommunication Standardization Bureau (TSB).
http://www.itu.int/en/ITU-T/techwatch/Pages/big-data-standards.aspx

Figure 2 - Patterns of adoption



[1] See appendix for detailed definitions and metrics used for value potential index.

Source: US Bureau of Labor Statistics; McKinseyGlobal Institute analysis (2011: 7)

By the same token, average M2M device penetration is around 2 percent of data subscriptions among measured networks, while it can reach 20 percent for those operators that focus on M2M. M2M communication represents a small share – around 0.1 percent – of total cellular traffic in terms of bytes. This traffic share will go up as LTE M2M devices and more powerful processors are included in high bandwidth and low latency-demanding applications such as consumer electronics, vehicles and billboards. However, the fast deployment of the new networks (4G, 5G[9]) will be key for the continuing growth of applications, to further unlock the development of all kinds of connected devices, including connected cars. The deployment and its pace cannot be taken for granted though as there is a noted gap between the explosion of traffic and usages and limited streams

[9] Still being standardized, the deployment is forecast for 2018-2020 introducing software defined 5G networks for "Anything as a service". Huawei predicts that 5G (10 Gbps) opens the "Internet of everything", bridging human non-human, virtual and physical communities: by 2020, 10 billion mobile terminal, 100 billion global wireless connections, lower latency of 1 millisecond. *Source:* http://www.mobileworldlive.com/category/huaweiart14

of revenues especially in the EU. Therefore, it is likely to be unevenly spread globally with some region like Asia [10] and the US [11] taking the lead.

According to the SAS 2013 Big Data Survey Research Brief, despite industry hype, most organizations have still to develop, implement or execute a big data strategy: only 12% of the surveyed organizations (SAS, 2013: 1). The same report concludes that while big data is a common theme is the market, organizations continue to be wary of its impact. A 2014 EY report noted as well that the big data revolution had not taken place so far for most companies: 63% of the French companies surveyed consider the notion was still too vague (EY, 2014: 27).

Earlier, the McKinsey report (2011) was acknowledging the same issue about the scope of implementation when stressing that a company's "data-driven mind-set" was to be a key indicator of big data's value to companies. Obviously, one cannot expect all companies to be data-savvy and to display this specific mind-set, which means that this is likely to become a barrier to entry or even to an appropriate implementation. This will require new expertise [12] and training. This may favour big companies as the costs to obtain data may be shrinking, but the expertise is expensive and not native to most industries. Besides, as noted in a White Paper on trends in digital book publishing, specialised technology suppliers and big data providers may favour deals with big companies, at least when the market is opening: "the IT vendors serving the publishing industry, including Klopotek [13] and Publishing Technology [14], focus on large international publishers" (McILROY, 2014: 8).

---

[10] China Mobile for instance is leading with 4G with already over 500,000 4G base stations (China Telecom only). The increased scale has also driven down the cost of 4G smartphones, which are now below $100 in China, and that has broadened the audience.
Source: http://www.mobileworldlive.com/category/huaweiart14

[11] In 2014, Google and Verizon announced that they are testing the capabilities for currently installed fiber networks to carry data even more efficiently – at 10 gigabits per second – to businesses that handle large amounts of internet traffic.

[12] The 2011 Mc Kinsey report was stating that 1.5 million more data-savvy managers were needed to take full advantage of big data in the United States.

[13] Klopotek is an international market leader in the area of publishing software, serving more than 350 publishers, and more than 14,000 users globally.
http://www.klopotek.com/en/homepage.htm.

[14] Publishing Technology is provider of content solutions that serves over 400 trade and scholarly publishers, including HarperCollins, McGraw-Hill, Macmillan, Elsevier, Springer, Sage, Oxford University Press, BMJ Group, Brill, United Nations, American Institute of Physics, American Society for Microbiology, BioOne and Bloomsbury Publishing.
http://www.publishingtechnology.com/about-us/

This is clearly why consultancies dealing with the issue are suggesting ways to initiate the transition. As DAVENPORT & TYCHE (2013:30) put it:

> "Organizations need to begin transitioning now to the new model" (a model they call Analytics 3.0). They add that "It means change in skills, leadership, organizational structures, technologies, and architectures", such a drastic change will require some time as the example of the "computerisation" of companies has illustrated [15].

■ The meaning of the phenomenon

Overcoming the barriers to a relevant implementation will create tensions among and within firms as long as the cost/benefit analysis remains uncertain. Therefore, the question is what this phenomenon means, or as noted by the US Executive Office Report on big data:

> "What really matters about big data is what it does" (United States Executive Office, 2014a: 9).

Neither the mere indication of the growth of the volume of data nor the more uncertain figures about the size of market(s) may be sufficient to understand the nature of the process.

Beyond the tensions we just highlighted, some are stressing the overall benefits it should bring. Although acknowledging some of the risks involved, the recent EC communication on "new strategy on big data" notes:

> "This global trend holds enormous potential in various fields, ranging from health, food security, climate and resource efficiency to energy, intelligent transport systems and smart cities, which Europe cannot afford to miss." (EC, 2014a).

This section gives some examples of sectors and companies that are implementing a big data approach. It opens up with the new players in this area, IT players and value added providers, then gives some examples of other industries.

Big data burst upon the scene in the first decade of the 21st century, and the first organizations to embrace it were online and start-up firms. Arguably,

---

[15] The literature on the topic is rather vast identifying the barriers, tensions; power struggles and the time-span involved.

firms like Google [16], eBay, LinkedIn, and Facebook [17] were built around big data from the beginning. They didn't have to reconcile or integrate big data with more traditional sources of data and the analytics performed upon them, because they didn't have those traditional forms.

Indeed IT players, labelled "the digital dragons" by Atelier Parisbas (2013:78), are intrinsically well placed to benefit from this shift, to make the most of new analytics, of big data and the cloud (Amazon together with Google [18] and Microsoft [19] is a leading provider of third party cloud computing services with AWS [20]). This shift brings along some marketing innovation. New digital players also pioneer the use of data mining for compiling viewers' recommendations (Amazon, Netflix, Pandora, Zynga, …).

Content providers are setting up entities to supervise the analytics. Amazon; LinkedIn created centres of excellence and Netflix fully centralised this activity. Netflix produced its hit show, *House of Cards*, after analysing the data from their consumer base (million "plays" per day, million searches, plus tags and other metadata) (CARR, 2013). In China, internet behemoths Baidu, Alibaba and Tencent are using big data "to change how entertainment is conceived, sold and consumed" (FRATER, 2015; JING, 2014). The videogames company Zynga relies on business unit organization, Google as well (BAIN, 2013a; van RIJMENAM, 2013). Zynga pioneered the use of big data by intimately linking game design and business models (Atelier Paribas, 2013: 81). Samsung uses big data to power the content recommendation engine on its newest smart TVs. No wonder that Atelier Parisbas (2013) holds that the big data phenomenon can be understood as the growing power of data in the economy of culture.

Box 2 - Value added content providers. Some cases.

- *Acxiom*, headquartered in Little Rock, Arkansas, founded as Demographics in 1969, is an enterprise data, analytics and software as well as a service company. It is one of the largest database marketing services and technology providers in the world. Acxiom collects and aggregates data from surveys, registrations, purchases, postings, etc.

16 Google File System is a proprietary distributed file system developed by Google; has been part of the inspiration for Hadoop.

17 Facebook developed Cassandra, an open source (free) database management system designed to handle huge amounts of data on a distributed system.

18 Big Query.

19 Azure.

20 Amazon created Dynamo another proprietary distributed data storage system.

- *AppAnnie* describes its mission as performing "The Math Behind the App Stores" through its technology infrastructure, analytics platforms, advanced statistical models and a world-class. Since it started operating in 2009, the company has tracked more than 60 billion downloads and 17 billion US $ of app store revenue across iOS and Google Play. Hence their claim of being one of the largest companies globally in terms of app store market data tracked. Its international team of over 240 people are working across Beijing, Hong Kong, Tokyo, San Francisco, London, Seoul, Moscow, Amsterdam, Shanghai and New York. App Annie is backed venture capital firms and have raised 39 US $ million in funding.

- *Criteo* processes a vast amount of rich purchase-intent data in real-time to identify buyers and deliver dynamically created ads which are personalized for each consumer. Criteo analyses 230 terabytes of data daily (representing 2 petabytes of raw uncompressed data). The Criteo Engine has been developed over the past eight years (2005-2013) and consists of multiple machine learning algorithms, in particular, prediction and recommendation algorithms. The accuracy of the prediction and recommendation algorithms is supposed to improve with every advertisement delivered, as they incorporate new data.

- *DataSift* claims to be the leading social data platform with more than 1,000 customers in over 40 countries. It was founded to help organizations improve their understanding and use of Social Media, to enable companies to easily aggregate, filter and extract useful data from the billions of public social conversations on social networks, blogs, news and other web sources. DataSift is focused on producing state-of-the-art data-filtering technology. DataSift started its life as an offshoot from TweetMeme, the highly popular Twitter news feed service.

- *Equifax* creates credit reporting data from many sources. Equifax generates 158 billion credit score updates per month and 60,000 updates per second for consumers to help them in their daily lives to buy houses, cars, finance educations, secure credit cards, purchase appliances and more. The company organizes, assimilates and analyses data on more than 600 million consumers and more than 80 million businesses worldwide and its databases include more than 200 million employee files. It has operations in the U.S. and 18 other countries.

- *Experian*, founded in 1996, in an Irish global information services group with operations in 40 countries, providing data and analytical tools, across many different markets, including financial services, direct-to-consumer, retail, telecommunications, automotive, insurance, media and technology, public sector and education, and healthcare payments. Total revenue for the year ended 31 March 2014 was 4.8 billion US $. Experian employs approximately 16,000 people in 39 countries and has its corporate headquarters in Dublin, Ireland, with operational headquarters in Nottingham, UK; California, US; and São Paulo, Brazil. Experian provides much of US Healthcare.gov's identity verification component using consumer credit information not available to the government.

- *Gnip*, founded in Boulder, Colorado in 2008, describes itself as a source for social data and claims to be the world's largest and "most trusted provider" of social data with customers in over 40 countries. They provide, for instance, managed public API access to a wide range of social data sources like Facebook, Instagram, Dailymotion, and YouTube.

- *Tarsus*, founded in 1998 and incorporated in 2006, has revenues of nearly £ 73 million and 322 employees, 51% of its revenues is coming from emerging markets. Tarsus is an international business-to-business media group with interests in exhibitions, conferences, education, publishing and online media. The company operates in three segments: Europe, USA and Emerging Markets. The company has six principal business sectors: aviation, medical, labels and packaging, discount clothing (Off-Price), France and online media.

Source: Companies website.

Besides, other "value added content providers" (IDC quoted by Forbes, 2014) (see Box 2) are often given as examples. These companies include "traditional vendors" such as Thompson LexisNexis (see Box 3, HPCC), and Experian; "new wave vendors" such as DataSift, Gnip, and LinkedIn; "company and personal information vendors" such as Acxiom, Equifax, and Tarsus. Specialised companies like App Annie, Distimo (bought by App Annie), and Flurry (bought by Yahoo) (see Box 4, Flurry) emerged over the last years and provide all kind of analysis and metrics to their customers.

Box 3 - High Performance Computing Cluster Systems (HPCC) at Reed Elsevier

The company is a technology oriented scientific and technical book publisher, investing 500 million US $ every year, claiming to be the fourth largest digital content provider in the world (Annual Report, 2013: 9) with global revenues of 7.5 billion euros.

High Performance Computing Cluster Systems (HPCC), HPCC Systems is one of the most advanced, fast-performing big data processing technologies available today according to the company. It was developed by LexisNexis Risk Solutions and currently powers core products from this division, which had 2013 revenues of £933 million. It is open source and used to solve large-scale, complex data and analytics challenges. HPCC Systems combines proven data processing methodologies with Reed Elsevier's proprietary linking algorithms

The latest version of SciVal, their scientific, technical & medical segment's tool for universities and other institutions to assess their relative performance, runs on HPCC Systems technology. SciVal provides analysis of over 30 million pieces of content and 350 million citations from 4,600 institutions in 220 countries. Lexis Advance provides lawyers with essential information and analytical tools covering all aspects of their daily work.

The company claims it offers their customers ways to turn data into intelligence – better, faster and cheaper.

Source: Annual Report, 2013.

Online platform providers process and analyse big data to find meaningful correlations in order to specifically target products and services to individual consumers. Platforms are poised to become an increasingly important point of aggregation in the advertising market with unique reach, customer data and measurability. By combining their traditional expertise with new techniques, including data analysis and viral marketing, publishers/ content aggregators can retain their valuable position as curators in the digital space and play a key role in ensuring quality content finds its audience. Big data will be used online and with mobile to power the content recommendation engine.

Box 4 – Flurry personalizing the mobile experience

Flurry, a US company headquartered in San Francisco, California, was founded in 2005 with the help of venture capital. Flurry describes its mission as to optimize the mobile experience through better apps and more personal ads. Flurry Analytics became the industry standard in mobile. Flurry is venture-backed and headquartered in San Francisco with offices in New York, London, Chicago and Mumbai. As of July 2014, the company was acquired by Yahoo.

- 170,000 developers use Flurry Analytics, Flurry works with mobile developers in 150 countries,
- Flurry sees app activity from 1.4 billion devices monthly,
- Flurry sees 5.5 billion app sessions per day,
- Flurry Analytics is in 7 apps per device on average,
- 8,000 publishers monetize with Flurry.

Flurry releases its mobile industry trends and insights on the Flurry Insights blog. Not being public, its revenues are not available but according to internet sources it became cash-flow positive in 2012.

Source: www.flurry.com, internet sources (KIM, 2012)

The reports from consultancies give examples of applications to other sectors like banking, retail, …To prove their case that "big data can play a significant economic role to the benefit not only of private commerce but also of national economies and their citizens", the authors of the McKinsey 2011 report concentrated on five domains: health care in the United States; public sector administration in the European Union; retail in the United States; global manufacturing; and global personal location data (McKinsey 2011: 37). They did conclude it was significant, even, as it noted earlier, (Table 1) unevenly distributed among sectors: for instance in the case of the European Union, they estimate that government administration could save more than €100 billion in operational efficiency improvements alone by using big data.

The recent High-level Expert Group report on "Taxation of the Digital Economy" also insists on the economic impact and its very significant proportions in the near future (EC 2014b: 51). By the same token, the press release announcing the creation of the Data PPP partnership, between the European Commission and the Big Data Value Association [21], gives the following figures for the EU: "up to 30% of the global data market for European suppliers; 100,000 new data-related jobs in Europe by 2020; 10% lower energy consumption, better health-care outcomes and more productive industrial machinery" (EC Press Release, 2014c).

[21] A non-profit, industry-led organisation.
http://www.bigdatavalue.eu/index.php/big_data_value_association

From another angle, DAVENPORT & TYCHE (2013) gave several examples of the use of big data by companies (Box 5: a sample) indicating the goal, the scope of the use and the technologies being used. The sample may not be representative but it does supply some information of what is being done, beyond IT players and "digital dragons", under the label "big data" within some companies, in that case US companies only.

Davenport and Tyche describe an initial era, they call "Analytics 1.0", during a period, which stretched 55 years from 1954 (when UPS initiated the first corporate analytics group) to about 2009. They claim that from 2005 to 2012, the era of "Analytics 2.0" began with the exploitation of online data in internet-based firms like Google, Yahoo, and eBay. According to the two authors, we are now moving toward an "Analytics 3.0" model that ushers in an emphasis on prescriptive analytics (DAVENPORT & TYCHE, 2014: pp.19, 26).

Box 5 - A sample of companies using of big data [22]

- *Bank of America* considers big data in three different "buckets"—big transactional data, data about customers, and unstructured data. Their primary emphasis is on the first two categories.

- *Blinkbox*, a former Tesco subsidiary sold to the TalkTalkgroup in 2015, is a movie streaming business using cloud service. The company created a SQL Server. Data are fed through a real time, stream processing framework where they are aggregated, analysed and then stored into a distributed database based on Hadoop stack technologies. The big data team uses technologies like Storm, HDFS, HBase, Hive, Impala, Pig, Sqoop.

- *Boo-box*, one of the top ad networks in Brazil, uses big data to hone its targeting and gain near real-time insights into the more than 3 billion ads it places on 350,000 blogs and websites each month. Under the Boo-box model, publishers earn money by hosting ads on their websites and blogs, while advertisers reach new customers by displaying ads on websites that deliver the best results.

- *Coursera* an educational UK company that collaborates with Stanford, Duke, Princeton, the London School of Economics and other institutions to offer free online classes (more than 300 classes on its website). Coursera can handle half a petabyte of traffic each month and scale to deliver courses over 3 million students in one year.

- *Diebold*, one of the largest global providers of automated teller machines, created a tablet-based application called Conductor that bank branch personnel use to monitor ATM usage. Conductor consists of a software agent that runs on the ATM and collects session data and feeds it to its cloud provider.

[22] BIG quotes the use of big data by Siemens. However there are no indication similar to the ones used in that box. It stresses selected growth fields: electrification, automation, and digitization. http://big-project.eu/sites/default/files/Value-of-EU-BigData-CSAs.pdf

- *GE* collects signals coming from sensors: "Our sensors collect signals on the health of blades on a gas turbine engine to show things like 'stress cracks.' The blade monitor can generate 500 gigabytes per day – and that's only one sensor on one turbine. There are 12,000 gas turbines in our fleet." GE has been the poster boy for this emerging trend, called "Analytics 3.0" by IIA, or "the digitization of everything" by Forbes.

- *Interactions Marketing*, founded in 1988, provides in-store product demonstrations and outdoor experience marketing programs for retailers and brands worldwide. With over 45,000 associates, the San Diego-based company creates and executes more than 2 million events every year and manages 5,500 events each day.

- *Macys.com* utilizes a variety of leading-edge technologies for big data; they include open-source tools like Hadoop, R, and Impala, as well as purchased software such as SAS, IBM DB2, Vertica, and Tableau.

- *redBus*, in 2006, the online travel agency introduced internet bus ticketing in India, unifying tens of thousands of bus schedules into a single booking operation (a kind of Expedia for bus booking.) redBus crunches terabytes of booking and inventory data in mere seconds and at a fraction of the cost of other big-data services.

- *Schneider National* (one of North America's largest truckload, logistics and intermodal services providers) builds on the availability of low-cost sensors for its trucks, trailers and intermodal containers. The sensors monitor location, driving behaviours, fuel levels and whether a trailer/container is loaded or empty.

- *Sears* is using big data technologies to accelerate the integration of petabytes of customer, product, sales, and campaign data in order to understand increased marketing returns and bring customers back into its stores. The retailer uses Hadoop and is now leveraging open source projects Apache Kafka and Storm to enable real-time processing.

- *UPS* tracks data on 16.3 million packages per day for 8.8 million customers, with an average of 39.5 million tracking requests from customers per day. The company stores over 16 petabytes of data.

- *United Healthcare* uses a variety of tools. The data initially goes into a "data lake" using Hadoop and NoSQL storage, so the data doesn't have to be normalized.

Source: Amazon-AWS, DAVENPORT & TYCHE (2013); FORBES (2014); Google-Big Query, Microsoft-Azure.

So far we just shown that new IT companies came up with new tools to sustain their growth building on a technological ground they were familiar with, seizing the opportunities enabled by technologies, bringing along various forms of innovation (in audience reach, in processes…). Besides, some companies from other sectors started introducing some technologies available to update and upgrade their approaches. However, the broader benefits that some sources are predicting or expecting may not be easy to collect. As noted by the World Economic Forum (WEF, 2013):

> "Fully tapping that potential holds much promise, and much risk". WEF rightly underlines that, "the economic and social value of big data does not come just from its quantity. It also comes from its quality – the ways in which individual bits of data can be interconnected to reveal new insights with the potential to transform business and society".

We already mentioned the (usual) barriers to changes. The next section will concentrate less on the opportunities, more on the issues stemming from the phenomenon, and considers the policies that may be required.

■ Big data/ big problems or big data/ big opportunities?

Tensions

The "digital dragons" act and will act as catalysts within the big data ecosystem, however at the same time the very rise of these "data barons" is triggering market concentration and data oligopolies issues. So far, as emphasized by Haire and Mayer, this trend of concentration has been countered by a lively ecosystem of big data start-ups, some of which succeed by positioning themselves well in the flow of information and compete well even against the very largest of big data companies. Certainly the fluidity of the big data ecosystem, enabled by low barriers to entry, enables these start-ups and acts as a counterforce to market concentrations. They nevertheless suggest to remain vigilant against what they hold as the "dark side of market concentration and data oligopolies" (HAIRE & MAYER, 2014: 18).

A related, although different issue, is the potential rise of new forms of conflicts of interests within a specific company growing a business out of big data: if data is getting too important it may conflict with other activities as happened in the case of banks with deposit branches and financial services. In the latter case it triggered new regulation. Data ownership and the legal ability to use the data is another linked issue especially as players coming from different industrial backgrounds have different regulatory frameworks dealing with this issue; what is possible is some sectors may not be authorized in others. Typically, telecom operators argue that the "digital dragons" can deal freely with data but as telecom operators are bound by stricter rules not to use the data for commercial purposes without customers' agreement, they cannot. Some broader more comprehensive regulation may

be needed as well, to create a "level playing field" to quote the usual industry buzz-word.

Big data raises concerns about the protection of privacy and other values as illustrated by a recent report for the second Obama administration (Podesta Privacy Report, 2014). Anderson and Rainie summed up the main results of a 2012 Pew Research Survey [23], stressing "our query rendered a decidedly split verdict", noting that "while enthusiasts see great potential for using big data, privacy advocates are worried as more and more data is collected about people" (ANDERSON & RAINIE, 2012:3).

A rethink of traditional approaches to data governance may be needed, particularly a shift from trying to control the data itself to focusing on the uses of data. Prevalent data standard protection may have become higher as legal standards are becoming more and more inadequate, individual notice and consent, opting out and anonymization may be losing out as protection modes.

Furthermore, it raised anew the question of how to balance intelligence and security (surveillance) without jeopardizing human rights. The nature of surveillance has changed according to MAYER-SCHÖNBERGER & CUKIER (2013). For instance, I. LEVY (GCHQ [24]) [25] deems that "anonymization is really hard, maybe impossible". As the director of the same UK agency, GCHQ, puts it, in any case one can never be totally safe from cyber threats and "the technical level of cyber-attacks is growing exponentially" (LOBBAN, 2013:15). These two issues (governance, security) illustrate the tensions that are inherent to the growth of this new industry and the difficult choices policy makers will be facing.

Predictive fallacies?

From a different social perspective, others fear that it might bring some type of "informational determinism", serving "the majority (at times inaccurately) while diminishing the minority and ignoring important outliers" (ANDERSON & RAINIE, 2012: 4). Besides, just spotting correlation may not

[23] Digital stakeholders were asked to weigh two contrasted scenarios for 2020, a positive and a negative one.

[24] GCHQ is a UK intelligence and security organisation.

[25] Presentation at the IIC Annual conference, Vienna, October 8-9, 2014. http://www.iicom.org/

suffice as a well-known issue for statisticians is that correlation is not causation "and sometimes, correlation is not correlation" further notes the Electronic Frontier Foundation in its comments to the US National Telecommunications and Information Administration (EFF, 2014). The "parable" of the Google Flu (GFT [26]) is often pushed forward as an example of such miscalculations (LAZER *et al*., 2014). Just deriving from Google or Facebook the assumption that people are the sum of their social relationships, online interactions and connections with content may just be simply wrong.

Besides, overoptimistic views may overlook that most likely lots of technical hurdles will have to be overcome. For instance, much of the data being generated now is "unstructured" and sloppily organized. Getting it into shape for analysis is no small task. From a more positive angle, DAVENPORT & TYCHE (2013: 30) stressed that:

> "It's important to remember that the primary value from big data comes not from the data in its raw form, but from the processing and analysis of it and the insights, products, and services that emerge from analysis".

Big data for development and marketing

At the same time, as underlined by R.Samarajiva, big data and data analytics may have the potential to improve the delivery of public services and thereby contribute to e-inclusion, specifically in the case of the developing-countries (SAMARAJIVA, 2014: 2). So as to foster the use of big data for development, the Executive Office of the United Nations Secretary-General launched, the UNPulse Initiative to mainstream the use of data mining and real-time data analytics into development organizations and communities of practice. Global Pulse is a flagship innovation initiative of the United Nations Secretary-General on big data, meant to promote awareness of the opportunities big data presents for relief and development.

We already hinted at the reduced barriers to entry as a feature of the big data ecosystem. This is a feature that may contribute to reducing any new digital divide coming from the use of big data or at least mitigate the divide by enabling more firms to enter. Indeed, big data start-ups do not have to

[26] To sum it up, GFT overestimated the prevalence of flu in the 2012-2013 season and overshot the actual level in 2011-2012 by more than 50%.

invest heavily in technical infrastructure to process and store data, they can utilize public cloud platforms (cloud storage and cloud processing capacity) that provide them with flexible commodity priced capabilities when they need it. This had not been the case with the previous generation of start-ups, such as Facebook and Google. One can assume that growth of data will create room for new business models. Some are already emerging based on three pillars: mobility/ broadband/ cloud (ERICSSON, 2014).

Also on the positive side, as the examples introduced in the preceding section show, big data has given birth to an array of new companies and has helped existing companies boost customer service and find new synergies (KAKUTANI, 2013). From a marketing viewpoint, the information that companies get about the consumption of their products is growing exponentially thanks to technology, allowing a better and highly sophisticated understanding of customers (WALLENSTEIN, 2013).

Enabling policies

Taking into account the potential benefits from big data, governments have started setting up policies. Part of the initiatives are designed to address the issue of the skills needed as reports and consultancies are pointing to the lack of expertise in the field. Prioritising data science in education and in training for early-career researchers is climbing up on the policy agenda. Other areas are being targeted like medicine with data-driven medicine [27].

In March 2012, the White House Office of Science and Technology Policy (OSTP) announced a Big Data Research and Development Initiative: six US government agencies were allocated over 200 million US$ to help the government better organize and analyse large volumes of digital data. Two years later the year, the US Executive Office released two reports [28] (US Executive Office, 2014a,b): one dealing with the discussion on privacy, the other adding a technology perspective. The latter concludes that "technology alone cannot protect privacy", and that policies are needed to protect privacy.

[27] See the adoption in the US in 2009 of the HITECH Act: taxpayers and the medical industry have collectively invested more than $100 billion in an information technology infrastructure.

[28] Prepared by the President's Council of Advisors on Science and Technology (PCAST).

As noted earlier, in July 2014, the European Commission outlined a new strategy on big data and launched €2.5 billion partnership to master big data. Besides, in Europe the "Big Data Public Private Forum" (BIG: see Box 6) was set up to contribute to the definition and implementation of a clear strategy to define the efforts in terms of research and innovation.

Box 6 - The "Big Data Public Private Forum" (BIG)

The BIG mission can be summarized as follows:

- Build a self-sustainable Industrial community around big data in Europe:
  - Technical level establishing the proper channel to gather information,
  - Industrial-led initiative to influence adequately the decision takers.
- Promote adoption of earlier waves of big data technology,
- Tackle adequately existing barriers as policy and regulation issues.

Outcomes of this 3 year project (started in 2012-2014), funded by the European Commission, will be used as input for Horizon 2020. The group will elaborate a sector roadmap.

Source: http://big-project.eu/

On both sides of the Atlantic, governements are dealing with the issues of data availability and owenership, of acces to data generated through public funds. The US National Science Foundation in the US and the EU Horizon 2020 framework programme contain provisions for open data management in research projects.

■ Conclusion

The growth of big data and its successful implementation relies on the deployment of networks with the proper "fit for big data" design as networks determine the user experience, it requires new devices [29] at affordable prices and appropriate device connectivity, and new platforms will be needed, as well as new techniques including in-memory analytics, and the right level of expertise; not an easy blend. One should therefore remain careful about hasty conclusions, big data may have arrived, but as stressed by HANFORD (2014): "big insights have not".

[29] For instance, with higher battery life to accommodate the data.

The World Economic Forum (WEF, 2013:3), stressing the need for a new approach, notes carefully that: "Fully tapping that potential holds much promise, and much risk", as the use of technology and data can both generate great value and create significant harm, sometimes simultaneously. There is no doubt about some kind of "zettabytes waltz" taking place, about a fast expansion of the digital universe. This digital universe is expanding and will continue to do so. However the pace, scope and locus of the expansion will vary. Some regions are already lagging behind, for example for the speed and bandwidth of networks, with Asia taking the lead for mobile broadband. Beyond the hype, some industrial sectors are better equipped than others to seize some of the opportunities.

In all cases it opens a lot of space for policymakers to intervene so as to create the right conditions for an implementation that could benefit all stakeholders. Governments have a role to play, not only to fine tune the regulatory framework, but also to make the most for the public sector, to pioneer the use of big data. These kinds of interventions should smooth out the transition to a productive use of big data. This transition will take time especially as education/ training is involved. Creating a "data-driven mind-set" so as to make up with the present imbalances is not an easy task.

Establishing an updated set of principles and the means to uphold them in a hyper-connected world is not easier either. Technology is part of the problem but can be part of the solution. Nevertheless, the transition will create tensions that may drive policymakers to put up some unnecessary hurdles. The balance between innovation and the protection of existing rights is always difficult to achieve as, if the goals are and remain legitimate, the ways to achieve them is likely to evolve.

For instance, notions of privacy change generationally, therefore trade-offs among privacy, security, and convenience will have to be based on a better understanding of what the nature of privacy is in the digital world, and to find those technological, educational, and policy avenues that will preserve and protect it, enable new industries to develop and new business plans to bloom.

References

Acxiom: http://www.acxiom.com/about-acxiom/

Akamai (2014): *The State of the Internet*. Q1 2014 Report/ Vol. 7/ N° 1. http://www.akamai.com/stateoftheinternet/

Amazon: http://aws.amazon.com/solutions/case-studies/

ANDERSON, J. & RAINIE, L. (2012): *The Future of Big Data, Pew Research*. http://www.pewinternet.org/2012/07/20/the-future-of-big-data/

Atelier Paribas (2013): *Big data, big culture? The Growing Power of the Data and its Outlook for the Economy of Culture*. http://www.forum-avignon.org/sites/default/files/editeur/EtudeATELIER_FA_2013.pdf

Bain (2013): *Big Data: The Organization Challenge*. www.bain.com

Boston Consulting Group (2012): *The Value of Our Digital Identity*, Liberty Global Policy Series. https://www.bcgperspectives.com/content/articles/digital_economy_consumer_insight_value_of_our_digital_identity/

BRYNJOLFSSON, E., HITT, L. M. & KIM, H. H. (2011): "Strength in numbers: How does data-driven decisionmaking affect firm performance?" April. ssrn.com/abstract=1819486.

CARR, D. (2013): "Giving Viewers What They Want", *New York Times*. http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?pagewanted=all&_r=0

CASTRO, D., (2014): "Center for Data Innovation, Comments to the National Telecommunications and Information Administration". http://www2.datainnovation.org/2014-ntia-big-data.pdf

CASTRO, D. & KORTE, T. (2014): "A Catalog of Every 'Harm' in the White House Big Data Report," The Center for Data Innovation, July 15, 2014. http://www.datainnovation.org/2014/07/a-catalog-of-every-harm-in-the-white-house-big-data-report/.

Cisco Visual Networking Index (2014): *Global Mobile Data Traffic Forecast Update, 2013-2018*, Februar. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

CLOUDERA (2014): *Le Guide du Big Data. 2014-2015*. http://bigdatainfo.fr/acces-au-telechargement.html

Criteo: http://www.criteo.com/about-us/

DataSift: http://datasift.com/company/

Davenport, T.H., Dyché, J., (2013): Big Data in Big Companies. International Institute for Analytics (iianalytics.com). Available at: http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf

Equifax: http://www.equifax.com/about-equifax/company-profile

EFF - Electronic Frontier Foundation, (2014): "EFF's comments to the National Telecommunications and Information Administration", https://www.eff.org/fr/document/comments-ntia-big-data

Ernst & Young (2014): *Etude Big Data 2014*. http://www.ey.com/FR/fr/Services/Advisory/EY-etude-Big-data-2014-strategie-big-data

EC - European Commission:

- (2014a): *Towards a thriving data-driven economy*. http://ec.europa.eu/digital-agenda/en/towards-thriving-data-driven-economy

- (2014b): High-level Expert Group Final report, *Taxation of the Digital Economy*, "Annex: Technological Trends". http://ec.europa.eu/taxation_customs/resources/documents/taxation/gen_info/good_governance_matters/digital/report_digital_economy.pdf

- (2014c): "European Commission and data industry launch €2.5 billion partnership to master big data". http://europa.eu/rapid/press-release_IP-14-1129_en.htm

e-skills UK/ SAS, (2013): "Big Data Analytics: An assessment of demand for labour and skills, 2012-2017".

EMC (2014): "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", EMC Digital Universe with Research & Analysis by IDC. http://www.emc.com/leadership/digital-universe/2014iview/index.htm

Ericsson Mobility Report (2014). *On the pulse of the networked society*. www.ericsson.com/ericsson-mobility-report

Experian: http://www.experianplc.com/

FORBES, G. (2013): "$16.1 Billion Big Data Market: 2014 Predictions From IDC And IIA". http://www.forbes.com/sites/gilpress/2013/12/12/16-1-billion-big-data-market-2014-predictions-from-idc-and-iia

FRATER, P. (2015): "China Rising: How Four Giants Are Revolutionizing the Film Industry", *Variety*. http://variety.com/2015/film/news/china-rising-quartet-of-middle-kingdom-conglomerates-revolutionizing-chinese-film-industry-1201421685/

Gnip: http://gnip.com/company/about/

Google: https://cloud.google.com/bigquery/

HAIRE, A. J. & MAYER-SCHÖNBERGER, V. (2014): "Big Data - Opportunity or Threat", ITU GSR discussion paper.

HARFORD, T. (2014): "Big data: are we making a big mistake?", *FT Magazine*. http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz3Jz6PNuSO

IDC:
- (2012): Worldwide Big Data Technology and Services, 2012–2015 Forecast.
- (2013): ITU's Technology Watch Report (November) 2013.

ITU (2013): http://www.itu.int/en/ITU-T/techwatch/Pages/big-data-standards.aspx

JING, L. (2014): "Big Data Fuels China's Film Industry Gold Rush", *Worldcrunch*, 16 Sept. http://www.worldcrunch.com/china-2.0/big-data-fuels-china-039-s-film-industry-gold-rush/alibaba-ebay-ecommerce-sohu-internet-cinema/c9s16906/

KAKUTANI, M. (2013): "Watched by the Web: Surveillance Is Reborn. http://www.nytimes.com/2013/06/11/books/big-data-by-viktor-mayer-schonberger-and-kenneth-cukier.html?pagewanted=all&_r=0

KIM, R. (2012): "Flurry rides the mobile boom, raises $25M with an eye toward IPO", 2 Nov. https://gigaom.com/2012/11/02/flurry-rides-the-mobile-boom-raises-25m-with-an-eye-toward-ipo/

LAZER, D, KENNEDY, R., KING, G. & VESPIGNANI, A. (2014): "The Parable of Google Flu: Traps in Big Data Analysis", *Science*, Vol. 343, n°6176, pp. 1203-1205.

LOBBAN, I. (2013): *Countering the cyber threat to business, IoD Big Picture*. http://www.gchq.gov.uk/press_and_media/news_and_features/Documents/directors_IoD_article.pdf

LANEY, D. (2001): "3D data management, controlling data volume, velocity and variety", Meta Group.

MAYER-SCHÖNBERGER, V. & CUKIER, K. (2013): A *Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin Harcourt.

McILROY, T. (2014): "11 Topmost Digital Book Publishing Trends & Opportunities". http://www.digitalbookworld.com/2014/eleven-digital-publishing-trends-for-2015/

McKinsey (2011): *Big Data: The next frontier for innovation, competition and productivity*. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

McKinsey Global Institute (2013): *Ten IT-enabled business trends for the decade ahead*. http://www.mckinsey.com/insights/high_tech_telecoms_internet/ten_it-enabled_business_trends_for_the_decade_ahead

Microsoft: http://azure.microsoft.com/fr-fr/case-studies/

PODESTA, J. *et al*. (2014): "Big Data: Seizing Opportunities, Preserving Values", Washington, Executive Office of the President, May 2014. http://www.whitehouse.gov/blog/2014/05/01/findings-big-data-and-privacy-working-group-review

Reed Elsevier (2014): *Annual Report 2013*. http://www.reedelsevier.com/mediacentre/pressreleases/2014/Pages/publication-of-annual-reports-and-financial-statements-2013.aspx

SAMARAJIVA, R. (2014): "The potential of 'big data' for improving public-service delivery in developing countries", LIRNEasia. http://lirneasia.net/search/

SAP (2014): "Beyond Connectivity. A Guidebook for Monetizing M2M in a Changing World".

SAS (2013): "2013 Big Data Survey Research Brief".
http://www.sas.com/resources/whitepaper/wp_58466.pdf

Tarsus: http://www.tarsus.com/about-us/about-us

TASCON, M. (ed.) (2013): "Big Data", *Telos*, n°95.

Transparency Market Research (2013): *Big Data Market - Global Scenario, Trends, Industry Analysis, Size, Share And Forecast 2012 – 2018*.
http://www.slideshare.net/wayne31jan/big-data-market-38557173

United States Executive Office:
- (2014a): *Big Data: Seizing Opportunities, Preserving Values*, Big Data Privacy Report.
http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf
- (2014b): *Big Data and Privacy: A Technology Perspective*.
http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

United Nations, UNPulse (2015): "Harnessing big data for development and humanitarian action". http://www.unglobalpulse.org/about-new

Van RIJMENAM, M. (2013): "Zynga: A Big Data Company Masquerading as a Gaming Company". http://smartdatacollective.com/bigdatastartups/116851/zynga-big-data-gaming-company

WALLENSTEIN, A. (2013): "Big Data: Media Embracing the Most Detailed Information About You Yet", *Variety*, September 25.
http://variety.com/2013/biz/news/big-data-media-embracing-the-most-detailed-information-about-you-yet-1200665847/

WEF - World Economic Forum (2013): "Unlocking the Value of Personal Data: From Collection to Usage", Prepared in collaboration with The Boston Consulting.Group.
http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf