7

Big Data—A Powerful New Resource for the Twenty-first Century

This chapter is a translation of an introductory article on "Big Data—Zauberstab und Rohstoff des 21. Jahrhunderts" published in Die Volkswirtschaft—Das Magazin für Wirtschaftspolitik (5/2014), see http://www.dievolkswirtschaft.ch/files/editions/201405/pdf/04_Helbing_DE.pdf. It reproduces the FuturICT blog of September 26, 2014, see http://futurict.blogspot.ch/2014/09/big-data-powerful-new-resource-for-21st.html, with minor improvements.

Information and communication technology (ICT) is the economic sector that is developing most rapidly in the USA and Asia and generates the greatest value added per employee. Big Data—the algorithmic discovery of hidden treasures in large data sets—creates new economic value. The development is increasingly understood as a new technological revolution. Europe could establish itself as Open Data pioneer and turn into a leading place in the area of information technologies.

When the social media portal WhatsApp with its 450 million users was recently sold to Facebook for $ 19 billion—almost half a billion dollars was made per employee. "Big Data" is changing our world. The term, coined more than 15 years ago, means data sets so big that one can no longer cope with them with standard computational methods. Big Data is increasingly referred to as the oil of the twenty-first century. To benefit from it, we must learn

D. Helbing, *Thinking Ahead—Essays on Big Data, Digital Revolution, and Participatory Market Society,* DOI 10.1007/978-3-319-15078-9_7

to "drill" and "refine" data, i.e. to transform them into useful information and knowledge. The global data volume doubles every 12 months. Therefore, in just two years, we produce as much data as in the entire history of humankind.

Tremendous amounts of data have been created by four technological innovations:

- the Internet, which enables our global communication,
- the World Wide Web, a network of globally accessible websites that evolved after the invention of hypertext protocol (HTTP) at CERN in Geneva,
- the emergence of social media such as Facebook, Google+, Whatsup, or Twitter, which have created social communication networks, and
- the emergence of the "Internet of Things", which also allows sensors and machines to connect to the Internet. Soon there will be more machines than human users in the Internet.

7.1 Data Sets Bigger than the Largest Library

Meanwhile, the data sets collected by companies such as ebay, Walmart or Facebook, reach the size of petabytes (1 million billion bytes)-100 times the information content of the largest library in the world: the U.S. Library of Congress. The mining of Big Data opens up entirely new possibilities for process optimization, the identification of interdependencies, and decision support. However, Big Data also comes with new challenges, which are often characterized by four criteria:

- *volume:* the file sizes and number of records are huge,
- *velocity:* the data evaluation has often to be done in real-time,

- *variety:* the data are often very heterogeneous and unstructured,
- *veracity:* the data are probably incomplete, not representative, and contain errors.

Therefore, one had to develop completely new algorithms: new computational methods. Because it is inefficient for Big Data processing to load all relevant data into a shared memory, the processing must take place locally, where the data reside, potentially on thousands of computers. This is accomplished with massively parallel computing approaches such as *MapReduce* or *Hadoop*. Big Data algorithms detect interesting interdependencies in the data ("correlations"), which may be of commercial value, for example, between weather and consumption or between health and credit risks. Today, even the prosecution of crime and terrorism is based on the analysis of large amounts of behavioral data.

7.2 What Do Applications Look Like?

Big Data applications are spreading like wildfire. They facilitate personalized offers, services and products. One of the greatest successes of Big Data is automatic speech recognition and processing. Apple's Siri understands you when asking for a Spanish restaurant, and Google Maps can lead you there. Google translate interprets foreign languages by comparing them with a huge collection of translated texts. IBM's Watson computer even understands human language. It can not only beat experienced quiz show players, but even take care of customer hotlines—often better than humans. IBM has just decided to invest $ 1 billion to further develop and commercialize the system.

Of course, Big Data play an important role in the financial sector. Approximately 70 % of all financial market transactions are now made by automated trading algorithms. In just a day, the entire money supply of the world is traded. So much money

also attracts organized crime. Therefore, financial transactions are scanned by Big Data algorithms for abnormalities to detect suspicious activities. The company blackrock uses a similar software, called "Aladdin", to successfully speculate with funds amounting approximately to the gross domestic product (GDP) of Europe.

To get an overview of the ICT trends, it is worthwhile to look at Google with its more than 50 software platforms. The company invests nearly $ 6 billion in research and development annually. Within just one year, Google has introduced self-driving cars, invested heavily in robotics, and started a Google Brain project, which intends to add intelligence to the Internet. By purchasing Nest Labs, Google has also invested $ 3.2 billion in the "Internet of Things". Furthermore, Google X has been reported to work on about 100 secret projects.

7.3 The Potentials Are Great...

No country today can afford to ignore the potentials of Big Data. The additional economic potential of Open Data alone—i.e. of data sets that are made available to everyone—is estimated by McKinsey to be 3000–5000 billion dollars globally each year [1]. This can benefit almost all sectors of society. For example, energy production and consumption can be better matched with "smart metering", and energy peaks can be avoided.[1] Resources can be managed more efficiently and the environment protected better. Risks can be better recognized and avoided, thereby reducing unintended consequences of decisions and identifying opportunities that would otherwise have been missed. Medicine can be better

[1] More generally, new information and communication technologies allow us to build "smart cities".

adapted to the patients, and disease prevention may become more important than curing diseases.

7.4 ... but also the Implicit Risks

Like all technologies, Big Data also imply risks. The security of digital communication has been undermined. Cyber crime, including data, identity and financial theft, quickly spread on ever greater dimensions. Critical infrastructures such as energy, financial and communication systems are threatened by cyber attacks. They could, in principle, be made dysfunctional for an extended period of time.

Moreover, while common Big Data algorithms are used to reveal optimization potentials, their results may be unreliable or may not reflect causal relationships. Therefore, a naive application of Big Data algorithms can easily lead to wrong conclusions. The error rate in classification problems (e.g. the distinction between "good" and "bad" risks) is often relevant. Issues such as wrong decisions or discrimination must be seriously considered. Therefore, one must find effective procedures for quality control. In this connection, universities will likely play an important role. One must also find effective mechanisms to protect privacy and the right of informational self-determination, for example, by applying the Personal Data Purse concept [2].

7.5 The Digital Revolution Creates an Urgency to Act

Information and communication technologies are going to change most of our traditional institutions: our educational system (personalized learning), science (Data Science), mobility (self-driving

cars), the transport of goods (drones), consumption (see amazon and ebay), production (3D printers), the health system (personalized medicine), politics (more transparency), and the entire economy (with co-producing consumers, so-called prosumers). Banks are losing more and more ground to algorithmic trading, Bitcoins, Paypal and Google Wallet. Moreover, a substantial part of the insurance business is now taking place in financial products such as credit default swaps. For the economic and social transformation into a "digital society", we may perhaps just have 20 years. This is an extremely short time period, considering that the planning and construction of a road often requires 30 years or more.

The above implies an urgent need for action on the technological, legal and socio-economic level. Already some years ago, the United States started a Big Data research initiative amounting to 200 million dollars; on top of this, further substantial investments were made. In Europe, the FuturICT project has developed concepts for the digital society within the context of the EU flagship competition. Other countries have already started to implement this concept. Japan, for example, has recently launched a $ 100 million 10-year project at the Tokyo Institute of Technology. Besides, numerous further projects exist, particularly in the military and security sector, which often have a multiple of the above-mentioned budgets.

7.6 Europe can Become a Motor of Innovation for the Digital Era

Europe can benefit a lot from the digital age. But it is not sufficient to reinvent and build already existing technologies in Europe. We must come up with new inventions, which will shape the digital age. The World Wide Web was once invented in Europe. The

largest civil Big Data competence in the world exists at CERN. Nevertheless, the USA and Asian countries so far have the lead in commercializing Big Data. With the NSA scandal, the spreading of wirelessly communicating sensors and the "Internet of Things", however, a new opportunity is emerging.

With a targeted support of ICT activities at its universities, Europe could take a lead in research and development. Switzerland, in particular, has academically excelled with the coordination of three out of six finalists of the EU flagship competition. At the moment, however, there is only a focus on the digital modeling of the human brain and robotics. From 2017 onwards, the ETH domain plans to increasingly invest into the area of *Data Science*, the emerging research field centered around the scientific analysis of data. However, in view of the fast development of the ICT area, the huge economic potential, and also the transformative power of these technologies, a prioritized, broad and substantial financial support is a matter of national interest.

With its democratic values, its legal framework, and its ICT engagement, including industry 4.0-related business, Europe might become an innovation motor for the digital age.

References

1. McKinsey & Company, Open data: unlocking innovation and performance with liquid information. http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information (2013). Accessed 28 April 2014
2. Y.-A. de Montjoye, E. Shmueli, S.S. Wang, A.S. Pentland, openPDS: protecting the privacy of metadata through SafeAnswers. PLoS ONE **9**(7), e98790. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0098790; see also http://newsoffice.mit.edu/2014/own-your-own-data-0709, http://www.taz.de/!131892/, and http://www.weforum.org/reports/personal-data-emergence-new-asset-class (2014). Accessed 28 April 2014