Is Big Data a Paradigm Challenge to Information Science?

Bruno Jacobfeuerborn
Deutsche Telekom AG, Germany

Abstract

Purpose: The purpose of the paper is twofold: (i) to argue that fundamental ideas are the most important tokens in scientific and engineering endeavours rather than specific methods, procedures, metrics and artefacts; new ideas and concepts are transformative forces of how we understand science, its role in society and how we lead scientific research; and (ii) to identify the challenges and opportunities the emerging concept of big data brings about to information science.
Approach/Methods: In order to investigate the impact of the new ideas in science we follow Thomas Kuhn's approach presented in his landmark book The Structure of Scientific Revolutions according to which science develops by leaps, which he dubbed paradigm leaps, that are qualitative changes of the ways the world is conceptualised and understood.
Results and conclusions: As a result of our investigation we identified four major paradigm leaps. The paper shortly depicts three paradigms that are already a canonical part of the past and contemporary science, and then a budding fourth paradigm that is still in statu nascendi, in its nascent stage, is described. We begin with Plato and Aristotle (first paradigm), and then through Francis Bacon (second paradigm), John von Neumann (third paradigm) we shall arrive at big data and knowledge discovery by means of computer facilities (potentially a fourth paradigm).
Originality/Value: It is believed that the fourth paradigm can help information technology become a partner on a par with humans in scientific and other research endeavours going far beyond its present role of being mainly a mechanism to store, process, and disseminate information. It is argued also that the fourth paradigm is a challenge to information science in both its main dimensions: (i) development of its foundations and methodologies by studying information phenomena reflected in very large datasets, and (ii) providing users with the needed information and knowledge derived from very large datasets.

Keywords Big data. Information science. Paradigms. Science.

Received: 4 November 2013. Revised: 7 November 2013. Accepted: 9 November 2013.

1. Introduction

From time to time we witness discussions and debates on the discrepancies between and the reasons why the Western civilisation that was born in ancient Greece, continued in Rome, redefined and consolidated in Renaissance and consecutive epochs in Europe, and then excelled in the United States outpaced and dramatically outperformed other great civilisations such as the Chinese or Arab ones in terms of material culture, human rights, standards of living and military capacities. This question is particularly thought-provoking because of at least two reasons. The first one is that after the glorious, flourishing, exploratory and revealing, and intellectually and socially vibrant times of Greece and Rome, over the

Middle Ages (legitimately dubbed Dark Ages by Petrarch or in Latin saeculum obscurum) of cultural and material deterioration, Europe dramatically lagged behind imperial China and Arabia in almost all respects, and all of a sudden as Phoenix arising from ashes it opened a new chapter of its history and started Rinascimento and begun an inexorable march to power and prosperity to its present apex and world domination with a hardly concealed pretension to universality and dictating political and economic norm and rules and championing its lifestyle and popular culture. The second reason is that having identified the actual sources and mechanisms of progress and prosperity we can attempt to devise and provide mechanisms to sustain the well-being and betterment as well as maintain a comparative advantage over those who are quickly catching up with Western standards of doing business, management, and living conditions. So, now the following question comes along: Which have been the causing factors of the Western comparative advantage? Most likely there is no definitive answer that might gain a consensus of pundits, historians, economist and sociologists, and politicians. Indeed, public discussions, popular and scholarly literatures offer an array of proposals ranging from simple or even simplistic to complex and sophisticated ones. Putting aside a survey of various theories and speculations on the Western ascendancy among which one can find a legacy of Roman Law, the role of Christianity, the separation of politics from religion ("the things of God are not those of Caesar"), and abandoning the dogma of divine sources of power and establishing a secular jurisdiction in a state, we tend to align our opinion with Niall Fergusson's conjecture of six headings: "What distinguishes the West from the Rest – the mainsprings of global power – were six identifiably novel complexes of institutions and associated ideas and behaviours. For the sake of simplicity, I summarise them under six headings: 1. Competition; 2. Science; 3. Property Rights; 4. Medicine; 5. The consumer society; 6. The work ethic. To use the language of today's computerised, synchronised world, these were the six killer applications – the killer apps that allowed a minority of mankind originating on the western edge of Eurasia to dominate the world for the better part of 500 years" (Ferguson, 2012).

Of these six mainsprings we particularly sympathise with the one that emphasises the role of science in creating conditions to boost, foster and award human creativity, intellectual curiosity and propensity to study, understand and explain nature leading to creating artefacts that make work more productive and efficient, and daily life easier, more pleasant and rewarding. We go even further and believe that fundamental ideas are the most important tokens in scientific and engineering endeavours rather than specific methods, procedures, metrics and artefacts. New ideas and concepts are transformative forces of how we understand science, its role in society and how we lead scientific research. As argued by Thomas Kuhn in his landmark book *The Structure of Scientific Revolutions* (Kuhn, 1996) science develops by leaps caused by new ideas, which are qualitative changes of the ways the world is conceptualised and understood. He dubbed such leaps paradigms.

The paper shortly depicts three paradigms that are already a canonical part of the past and contemporary science, and then a potential fourth paradigm that is still in *statu nascendi,* in its nascent stage, is described. We begin with Plato and Aristotle (first paradigm), and then through Francis Bacon (second paradigm), John von Neumann (third paradigm) we shall arrive at big data and knowledge discovery by means of computer facilities (potential fourth paradigm). The paper will address some questions and challenges to information science in the light of the fourth paradigm emergence.

Before we start the discussion it is good to remind that the terms "science" and "scientist" appeared in the colloquial language only in the 19th century; past that the people who did what now we call research considered themselves natural philosophers. However, in this paper we shall make use of these contemporary terms also while discussing the two paradigms that took place before the 19th century.

2. Four Paradigms

Aristotle's seminal work entitled *Organon* (Aristotele, 2009) is a collection of texts that laid out foundations for what is now called standard logic and till now is a common tool for carrying out scientific endeavours. Aristotle was Plato's favourite pupil who shared his love to philosophical ponderings and knack for precision thinking and reasoning, yet their conceptualisations of the world and methodologies to describe, analyse and interpret the world diverged at a certain point of their work and relationship. Incidentally, it is rather a classic case among intellectual giants that at a certain moment a disciple contests his master and comes out with a competitive approach. Plato's philosophy was based on an idealistic assumption whose essence was that the truth and prototypes of entities that we can see and experience in our surroundings lay in the ideal, transcendent space that is not reachable for humans who can through their senses see and feel only the shadows, incomplete and imperfect incarnations of the timeless and absolute original forms existing in this ideal external and eternal universe. He argued that all the attempts to comprehend these ideal objects by dissecting and examining their physical counterparts that are available to us is nothing else than an act of intellectual impudence and audaciousness that in all probability would lead to false claims. Therefore, a contemplation and philosophical insight and thoughtfulness, namely the engagement of the mind without touching physical objects in the process of tackling and understanding things and processes, are the right methodological tools to learn the truth. This ascetic and seeking perfection approach allowed deduction as the only inference mechanism to proceed from premises accepted beforehand to conclusions, for this was the sole reasoning method that guaranteed that given true premises the obtained conclusions must be true by virtue of the deduction itself. Incidentally, these were the ground assumptions of Euclid's geometry that is an intellectual masterpiece and a prefiguration of later axiomatic theories. Although Aristotle did not share Plato's idealistic ontology he also, as his mentor, cherished precision and certainty. One can guess that when Aristotle begun his study on human reasoning he looked for reliable and infallible principles that dependably governed the ways on how people think and express themselves in words (conversations, speeches and the like), the principles that would be so universal, necessary, and timeless as the rules that unvaryingly determine physical phenomena. This is how he discovered such basic logical principles as the principle of identity ("a thing is the thing it is"), the principle of uniqueness ("no thing is another thing than the thing it is"), or the principle of excluded middle ("a thing has or does not have a particular property"), and eventually established a framework and foundations for standard logic. Let us here put forward a hypothesis that the appreciation and even the insistence on reliable ways of reasoning were caused not only by the personas of Plato, Aristotle and other ancient thinkers, but also by the fact that past them individual and generalised experiences were orally transmitted through generations, which process was obviously prone to distortions and in many cases misleading information, and therefore prompted Aristotle to make it as reliable and verifiable as possible. What was said so far is the characterisation of the *first paradigm*, which is still a viable, recognised and widely practiced approach by contemporary scientists, scholars, and scientific projects; noticeably the paragon incarnation of this paradigm is the Euclidian geometry mentioned above. For the second paradigm one had to wait some twenty centuries.

Completeness, elegance and a pervasive and persuasive appeal of the first paradigm, additionally firmly supported by the Middle Ages Christian zeitgeist of absolutism and the division of the universe into two separate kingdoms, namely God's ideal realm and the imperfect human world on the planet Earth, discouraged thinkers not only to question the first paradigm based on Platonic idealism and Aristotle's natural philosophy and logic, but also to further elaborate on it towards modifications, enhancements or even to come out with alternatives. The fresh current of independent thoughts and makings, a new way of explaining natural phenomena referring to the best ancient traditions of Greece and Rome made its particularly noticeable appearance in Italian Florence around the 14th century, under the patronage and incentives of the Medici. Also in other Italian city-states (such as Bologna, Genoa, Milan, Padua, Siena, Turin, Venice, Verona) the spirit of Renaissance occurred. The fruits of this movement were the works by such geniuses as Michelangelo, Leonardo da Vinci, Francesco Petrarka, Niccolo Machiavelli, Girolamo Cardano, Galileo Galilei and many others. A sort of manifesto and summary of the Italian *Rinascimento* was pronounced by Pico della Mirandola in his famous and influential public discourse *Oration on the Dignity of Man* in the year of 1486 (Pico della Mirandola, 1994) . As Copernicus moved the centrality of the universe from the Earth to the Sun, likewise the Renaissance shifted away the interest of thinkers, pundits, physicians, travellers and adventurists, and artists from the transcendent world of God to the world of people. That was well reflected in the Renaissance mantra *Homo sum, humani nihil a me alienum puto* (I am human, and nothing human is alien to me). Thus the process of what Max Weber later dubbed *Entzauberung der Welt* (disenchantment of the world) begun (Weber, 1919). In 1543 Andreas Vesalius, a Brabantian physician and anatomist, published his seven-volume ground-breaking book entitled *De humani corporis fabrica* (On the Fabric of the Human Body) (Vesalius, 1543–1555) that by questioning the previously dogmas of truths paved the way to medical research based on mechanistic view of anatomy, the crucial role of dissection, and personal experimentations and observations. In the same year, just before Nicolaus Copernicus' death his revolutionary book that put upside down the cosmology entitled *De revolutionibus orbium coelestium* (On the Revolutions of the Heavenly Spheres) (Copernicus, 1543) was published. These two landmark oeuvres proved the rightness and fruitfulness of empiric approach to studying nature and carrying out scientific endeavours. Incidentally, this methodology was significantly advanced by inventing a telescope by two Dutchmen Hans Lippershey and Zacharias Janssen in 1608 and considerably improved by Galileo in 1609, and by inventing a microscope whose invention is also credited to Galileo in 1610 who called it *occhiolino*.

The spark of Italian *Rinascimento* quickly became a flame that spread throughout Europe and through geographic discoveries, Gutenberg's invention of the printing press and through the Protestant Reformation led directly to the Enlightenment, the Age of Reason, by promoting and championing the slogan of *sapere aude, incipe* (dare to know and to begin) coming from the Roman poet Horace and later picked up by Immanuel Kant in order to oppose to superstitions, dogmas, absolutism, intolerance, obstructive traditions and self-imposed constraints. In his landmark book *Was ist Aufklärung?* (*What is the Enlightenment?*) Immanuel Kant wrote: "Enlightenment is man's emergence from his self-imposed nonage. Nonage is the inability to use one's own understanding without another's guidance. This nonage is self-imposed if its cause lies not in lack of understanding but in indecision and lack of courage to use one's own mind without another's guidance. Dare to know! (Sapere aude.) 'Have the courage to use your own understanding', is therefore the motto of the Enlightenment" (Kant, 1996).

The Enlightenment and scientific revolution, following and enhancing the trends initiated in the Renaissance, put particular emphasis on collecting raw real-world data, as a first stage of scientific investigation, which dramatically changed the way of tackling problems and looking for their solutions. This is exactly how natural sciences such as biology, chemistry, geology, physics and other disciplines have systematically approached issues of their interest since the16th and 17th century. Francis Bacon's new methodology of science and knowledge, *empiricism*, that relayed on observation, collection of data, and experimenting, along with accepting induction as a legal inference method for scientific endeavours can be characterized as data-centric. Indeed, innate concepts, *a priori* assertions based on tradition, intuition or revelations could not be accepted as knowledge until they were verified and confirmed by rigorously organised experiments and the data the experiments yielded. Baconian science posits that theories that are meant to be the models of reality are derived from the analysis and generalisation of the collected data and observations, or if the models are established as intellectual hypothesis, they must be verified through experiments producing data that in turn have to be examined and scrutinized. In his seminal book entitled *New Organon or True Directions Concerning the Interpretation of Nature* (Bacon, 1620) whose title makes a clear reference to Aristote's *Organon* Bacon presented this new approach based on a pragmatic vision of the world and the assumption that humans are capable to reveal and understand the mechanisms of nature by putting at work their intellectual faculties and capacities to operate on the collected real-world data without referring to a divine prompting. Allowing one to experiment and collect data and use induction for setting up hypotheses was obviously a different methodology from the Aristotelian one, but at the same time the new approach did not falsify the first paradigm; it simply considerably enhanced it. Therefore, Bacon could legitimately entitled his oeuvre New *Organon*. This Baconian paradigm is nowadays applied not only in the scientific realm but also in a modified and less rigorous form in other domains such as marketing, politics or governance, for instance to learn social preferences and moods. This Baconian methodology that we consider a *second great paradigm* of science gained wide acceptance and was usher into practice and enhanced and strengthened the attitude of disenchanting the world in the Weberian sense.

One of the major concerns of scientists and inventors who followed the Baconian approach was that usually they either did not have enough data to draw conclusions and build up models, or on the contrary, there was too much data for a man or even a team to grasp it and discover patterns and regularities. The former problem has gradually been solved with time thanks to the advancements in laboratory equipment, sensors, and measurement instruments to collect and store data. In order to resolve the latter issue researchers waited for the appearance of computers organised according to the principles and architecture proposed by John von Neumann, i.e. until the mid of the 20th century, especially for large computer systems to crunch bulky amounts of data. At the beginnings of the computer era the ability to process large datasets caused indeed a quantitative change in dealing with empirical data, which was a considerable step forward but not a real qualitative breakthrough factor. At this point one should mention database management systems (DBMS), which was a major technology that was developed for structuring, storing, and processing of numeric and nonnumeric large datasets, and that offered user-friendly retrieval languages based on the SQL (Structured Query Language) approach. The revolutionary change, a significant and dramatic qualitative leap occurred when computers were engaged for simulation and modelling of physical and social phenomena, and to assist in mathematical studies. For the sake of periodization we assume that when Kenneth Appel and Wolfgang Haken proved the famous four-colour theorem1 in 1976 by means of extensive computations the *third paradigm* in science was born. An important recognition of the significance of computer modelling is the Nobel Prize in chemistry awarded in 2013 to Martin Karplus, Michael Levitt, and Arieh Warshel for "The development of multi-scale models for complex chemical systems", the work that was begun in the 1970s. By means of a system of computer programs they conceived, designed and developed they modelled chemical reactions that occurred at a very high speed, at fractions of milliseconds with which classical chemistry observations could not keep up. Another vital and sophisticated case of computer modelling and simulation is computer-aided drug design where robust computing facilities are used while and for conceiving, designing and studying new medications *in silico*, i.e. in computer environments and frameworks, rather than in bio – and chemical laboratories, or better to say, prior the drugs will be subject to tests *in vitro* and *in vivo*.

The last decade of the previous century brought about an interesting new direction in computer research and applications, namely data and text mining, which are techniques that strive for, putting it in a nutshell, transforming data into knowledge by extracting rules, regularities and patterns that are supposedly hidden in datasets. This technique has turned out particularly fruitful and productive in these areas where large datasets are available as a result of routine business activities, and a deep analysis of data aimed at discovering the knowledge these datasets conceal is required. This takes place for instance for marketing purposes to better understand customers' preferences and habits, for market basket analysis, for telecommunications networks to discover traffic anomalies, or for scientific research in human genetics. It has to be noted that over the last years various and numerous computer applications throughout the world based on networks, sensors, social activities, and divers cloud computing facilities have been generating an enormous amounts of data every second, for example an experiment led at the Large Hadron Collider at CERN generates some 40 Terabytes of data, every 30 minutes of a jet flight yields about 10 Terabytes of data, Google receives some 2 million search queries a minute, more than 294 billion email messages are sent a day; such breath-taking examples can go on and on. In the year of 2010

[1] "In mathematics, the four colour theorem, or the four colour map theorem, states that, given any separation of a plane into contiguous regions, producing a figure called a map, no more than four colours are required to colour the regions of the map so that no two adjacent regions have the same colour" (*Four-colour Theorem,* 2013).

"The Economist" magazine estimated that the mankind created 1,200 Exabytes of data, and the IDC, consultancy, estimates that the digital universe will reach 40 Exabytes in the year of 2020 (Ganz & Reinsel, 2012). A bibliometric analysis of publishing activities in the world proved that in the year of 2012 some 1.57 million scientific journal articles appeared, meaning 3 new papers per minute were published (Ferstein, 2013). This phenomenon of very large datasets has been dubbed "big data". It has turned out however that the available methodologies and tools of the currently existing database management systems cannot cope with big data, its size and velocity of growth. Thus, a new trend in computer science, and computer and software engineering to master big data and to tap into these very large pools of datasets has recently emerged. In (Ganz & Reinsel, 2011) emphasis is made on the fact that big data is not a "thing", it is rather an activity that benefits from various ICT methodologies and technologies: "Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis." A more profound discussion on the big data as a game changer methodology and technology whose major potential is in discovering hidden value and knowledge in immense datasets in various areas of human activities, in particular in science and governance is published in (Jacobfeuerborn & Muraszkiewicz, 2012), and reflections on the relationship between data, information, and knowledge can be found in (Jacobfeuerborn, 2013).

In a controversial paper entitled *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* Chris Anderson asked the following astonishing question: "What can science learn from Google?" and provided the readers with the following answer: "We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot" (Anderson, 2008). This audacious yet perky statement is in fact a definition of the new approach and methodology in carrying out scientific research. It is indeed the idea of the *fourth paradigm* in science. For the sake of historical accuracy let us note that this pioneering idea was for the first time articulated by Jim Gray during his talk to the Computer Science and Telecommunications Board, Committee on National Statistics, in Mountain View in California on January 11, 2007 where he said: "I wanted to point out that almost everything about science is changing because of the impact of information technology. Experimental, theoretical, and computational science are all being affected by the data deluge, and a fourth, 'data intensive' science paradigm is emerging. The goal is to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other. Lots of new tools are needed to make this happen" (Gray, 2009). Incidentally, the book *The Fourth Paradigm: Data-Intensive Scientific Discovery* in which the article cited above has been published, includes a collection of essays that thoroughly elaborate on Gray's idea.

3. A Challenge to Information Science

Information science is a unique discipline that in its vast multidisciplinary body includes disciplines whose focus is on information and knowledge. It deals with a large spectrum of problems among which are the philosophy of information, information architecture and knowledge organisation, ontologies, information ecology, information retrieval and languages for information seeking, information management, information needs analysis, information acquisition and dissemination, and – last but not least – bibliometrics. It is tightly related to librarianship and library science, archivistics, museology, computer science, linguistics, cognitive sciences, and studies on information and knowledge society. Information science has been developing its own methodology, subject to constant progress and change, which is focused on methods of knowledge acquisition, representation and classification, on textual and multimedia objects retrieval, on semantic analysis of text, and on users' needs analysis, to mention just a few topics out of many. Information science is a horizontal approach that traverses across an array of sciences freely borrowing methodologies from them. On the other side it covers a vast area of interest that is exploring, conceptualising, and evaluating the realm of information on sciences and their outputs such as publications or conferences. In practical terms it consists, among others, in collecting, classifying, clustering, storing, retrieving, aggregating and disseminating information on sciences or disciplines it takes into account by means of a wide range of media starting with a word of mouth, to catalogues, to printed and/or electronic newsletters and bulletins, to personalized information updates, to running specialized portals, and to bibliometric analysis and reports. This is how information science supports researchers, scientific communities, and also practitioners operating in administration, business, education, health and other domains of life.

We can sum up the above as follows: Information science has two dimensions, namely (i) it implements and practices an interdisciplinary approach to the research of its own foundations and methodologies; and (ii) it helps researchers, scholars, engineers, inventors, and other knowledge workers to locate and acquire information that is necessary in their works. Now, we can ask whether there is any role to play for big data and the fourth paradigm in the realm of information science? It goes without saying that as far as the first dimension is concerned big data can help find solutions of the problems that are subject to a given on-going research. Here, the challenge is exactly the same as it stands before the big data approach in general, i.e. to mange and cope with immense datasets that can grow at a very fast pace. Incidentally, at this point we should mention a new trend in information science that was dubbed "data science" that is a collection of various methodologies and practical approaches whose main objective is to derive meaning and value from very large datasets, in other words, the ambition of data science is to master big data. The second dimension is about serving users and providing them with the needed information and knowledge. Should the users know their needs the challenge is again the same as to any typical use of big data. But often we are faced with the situation of the sort "we do not know, what we do not know", meaning we are not aware of our ignorance. Thus we are not able to ask appropriate questions to realise the area of ignorance and perhaps to fill out the gaps existing in the corps of our knowledge. Parenthetically, the faculty to ask good questions, especially in science but also in business, may be more appreciated than to find solutions – sometimes questions may be more important than answers, as the latter sooner or later can be found. This is here where there is a vital role to play for big data since through the methodology of the fourth paradigm the areas of ignorance can be identified and brought about to user's attention, and thereby big data can help ask apt and astute queries and define fitting heuristics. This could be a new role of scientific information that could not only be a mechanism to look for information and answer questions but also to become a partner of humans in scientific and other research endeavours. To get implemented this idea and incorporate it in information science routine practices is a true challenge to face with, but at the same time it is a rare opportunity to provide it with new methodologies and tools, and to identify and conquer new territories for scientific investigations.

4. Final Remarks

Contemporary people, fascinated and somewhat blinded by spectacular achievements of science and technology over the last two centuries, or so, tend to forget or neglect that Western science has its long history going back to ancient Greece, and that their recent tremendous successes, accomplishments, and attainments are a result of a long accumulative process in which disruptive ideas, new paradigms and intellectual prowess have played a decisive role. From a bunch of theories on how science develops, in this paper we stuck to Thomas Kuhn's theory of scientific paradigms, which are the ways of viewing, dissecting, understanding and analysing subjects to scientific research, which dramatically change or even undermine existing theories, patterns and rules that are generally recognised valid and practiced by scientific communities at the moment when the new paradigm comes along. Evolution whose mechanism works in small steps rather than in big leaps, operates locally and contributes to improvements, refinements and "linear" enhancements within an existing scientific paradigm; therefore it is not an appropriate model to depict the progress of science. Noteworthy, these days also these small steps are sometimes called paradigmatic by scientists, science writers or commentators, for instance, a shift from Codd's relation model of data to object-oriented model of organising data structures happens to be referred as a paradigmatic shift in the field of database management systems. Perhaps a proclivity to practice this habit is entailed by the need for valorising or emphasising a particular new method or solution and exposing its novelty. We do not share this inclination and conservatively preserve the term "paradigm" for an actual disruptive change of a scientific methodology, which is so deep that it horizontally affects various scientific domains, going across different fields and specialisations. Should the adherents of a more liberal usage of terminology for whom paradigm shifts are just vital changes within a given area of science doggedly stick to their habits, we are ready to compromise on terminology and to substitute the term "meta-paradigm" for "paradigm" in our discourse.

At the end of this paper we cannot avoid mentioning the existence of scepticism and criticism regarding big data and its paradigmatic potential. D. Boyd and K. Crawford define big data as "a cultural, technological, and scholarly phenomenon that rests on the interplay of:

(1) Technology: maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets;
(2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims;
(3) 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" (Boyd & Crawford, 2012, p. 663).

The third bullet of this definition draws attention to the fact that in certain circles big data has become a mythology or a slogan that carries the promise of finding a Holy Grail for pursuing a total automation of research and knowledge discovery by means of computers. Such naïve opinions could lead to the conviction that a delegation of scientific research from humans to computers will be possible and eventually will take place, and *per analogiam* to Fukuyama's end of history (Fukuyama, 2006), it will give rise to the end of science as it has been done so far. Nothing of this kind of opinion agrees with our position that can be summarised as follows: If the fourth paradigm becomes a real thing, science as such, and routine scientific research will be functioning as a blossoming partnership of humans and computers equipped with interactive facilities for deep analysis, synthesis and discovery of knowledge.

Acknowledgments

The author wishes to thank Prof M. Muraszkiewicz of the Warsaw University of Technology for a valuable discussion regarding the role computers play in scientific research, especially for modelling and simulation, and for the hint about the pivotal importance of questions in science.

References

Anderson, Ch. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete [online]. *Wired Magazine*, 23 June [29.10.2013], http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Aristotle (2009). *Organon*. From 1a to 164 a according to Bekker numbers [online]. Transl. under the editorship of W.D. Ross. Internet Archive [29.10.2013], http://archive.org/stream/Aristotle-Organon/AristotleOrganoncollectedWorks_djvu.txt

Bacon, F. (1620). *The New Organon or True Directions Concerning the Interpretation of Nature* [online]. Liberty Library of Constitutional Classics [29.10.2013], http://www.constitution.org/bacon/nov_org.htm

Boyd, D., Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, vol. 15 (5), 662–679.

Copernicus, N. (1543). *De revolutionibus orbium coelestium* [On the Revolutions of the Heavenly Spheres] [online]. ADS Digital Library [29.10.2013], http://ads.harvard.edu/books/1543droc.book/

Fergusson, N. (2012). *Civilization: The West and the Rest*. Penguin Books

Firstein, S. (2013). *The Pursuit of Ignorance* [online]. TED. Talks [7.10.2013], http://www.ted.com/talks/stuart_firestein_the_pursuit_of_ignorance.html

Four-colour Theorem (2013). In: Wikipedia. The Free Encyclopedia [29.10.2013], http://en.wikipedia.org/wiki/Four_color_theorem

Fukuyama, F. (2006). *The End of History and the Last Man*. New York: Free Press; Reprint edition

Ganz, J., Reinsel D. (2011). *Extracting Value from Chaos* [online]. IDC Report [29.03.2012], http://idcdocserv.com/1142

Ganz, J., Reinsel D. (2012). The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East [online]. IDC iView [29.10.2013], http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

Gray, J. (2009). A Transformed Scientific Method. In: T. Hey, S. Tansley, K. Tolle (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft Corporation.

Jacobfeuerborn, B. Muraszkiewicz, M. (2012) ICT and Big Data as a Game Changer. In: Z.E. Zieliński (ed.) *Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne*. Kielce: Wydaw. Wyższej Szkoły Handlowej, 51–60.

Jacobfeuerborn, B. (2013). Reflections on Data, Information and Knowledge. *Studia Informatica* vol. 34 (2A), 7–21.

Kant, I. (1996). *What is Enlightenment?*[online] Transl. by M. C. Smith. Columbia University. CC Required Readings [29.10.2013], http://www.columbia.edu/acis/ets/CCREAD/etscc/kant.html

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*. University of Chicago Press, 3rd edition.

Pico della Mirandola, G. (1994). *Oration on the Dignity of Man* [online]. University of Michigan. Cosma's home Page [29.10.2013], http://vserver1.cscs.lsa.umich.edu/~crshalizi/Mirandola/

Weber, M (1919). *Wissenschaft als Beruf*. München und Leipzig Verlag von Duncker & Humblot [online]. Wikosource, German version [29.10.2013], http://de.wikisource.org/wiki/Wissenschaft_als_Beruf

Vesalius, A. (1543-1555). *De humani corporis fabrica. On the Fabric of the Human Body* [online]. Book One. An annotated translation of the 1543 and 1555 editions of Andreas Vesalius' De Humani Corporis Fabrica by D. Garrison and M. Hast. Historical introduction by Vivian Nutton. Northwestern University Evanston, IL USA [9.11.2013], http://vesalius.northwestern.edu/flash.html

---

Czy Big Data jest paradygmatycznym wyzwaniem dla nauki o informacji?

Abstrakt

Cel/Teza: Cel artykułu jest dwojaki: (i) uzasadnienie tezy, że podstawowe idee są ważniejsze w kształtowaniu przedsięwzięć naukowych i inżynierskich niż określone metody, procedury, miary i artefakty; nowe idee i pojęcia są siłami transformacji naszego rozumienia nauki, jej roli w społeczeństwie i sposobów prowadzenia badań; (ii) wskazanie wyzwań i możliwości, które dla nauki o informacji niesie rozwijające się pojęcie „big data".

Approach/Methods: Badając oddziaływanie nowych idei w nauce przyjęliśmy podejście Thomasa Kuhna przedstawione w słynnej książce Struktura rewolucji naukowych, zgodnie z którym nauka rozwija się cyklami, nazwanymi przez niego paradygmatami, które jakościowo zmieniają sposób konceptualizacji i rozumienia świata.

Results and conclusions: Rezultatem badań jest wskazanie czterech głównych paradygmatów. W artykule krótko przedstawiono trzy paradygmaty, które dziś stanowią kanoniczną część przeszłej i współczesnej nauki, a następnie opisano obiecujący paradygmat czwarty, pozostający jeszcze *in statu nascendi*, w fazie kształtowania się. Rozpoczęliśmy od Platona and Arystotelesa (pierwszy paradygmat), następnie poprzez Francisa Bacona (drugi paradygmat), Johna von Neumanna (trzeci paradygmat), dotarliśmy do „big data" i odkrywania wiedzy za pomocą narzędzi komputerowych (potencjalnie czwarty paradygmat).

Originality/Value Wyrażono przekonanie, że czwarty paradygmat może pomóc w przekształceniu technologii informacyjnej w równorzędnego partnera ludzi w przedsięwzięciach naukowych i innych zamierzeniach badawczych, znacznie wykraczającego poza jej obecną rolę głównie mechanizmu przechowywania, przetwarzania i rozpowszechniania informacji. Uzasadniono także opinię, że czwarty paradygmat stanowi wyzwanie dla nauki o informacji w obu jej podstawowych wymiarach; (i) rozwijania jej podstaw teoretycznych i metodologicznych przez badanie zjawisk informacyjnych uwidaczniających się w wielkich zbiorach danych, i (ii) zapewnianiu użytkownikom potrzebnej im informacji i wiedzy derywowanej z wielkich zbiorów danych.

Keywords Big data. Nauka. Nauka o informacji. Paradygmat.

Dr BRUNO JACOBFEUERBORN is a Director of Technology at Telekom Deutschland GmbH and Member of its Board, and Chief Technology Officer (CTO) of Deutsche Telekom AG. Over the period of 2007–2009 he was the Board Member and Director of Technology, IT, and Procurement of Polska Telefonia Cyfrowa (now T-Mobile Polska). In parallel with his professional activities qua manager, promoter and implementor of innovative solutions he carries out research in the areas of: models of innovation in high-tech environments, knowledge and information organisation to boost innovativeness, and impact of ICT technology on education and society. Towards this end, he has been cooperating with the Warsaw University of Technology, in particular with the BRAMA Lab that is an incubator of innovative mobile applications and facilities involving students and young researchers, and with the Institute of Information Science and Book Studies of University of Warsaw, where he received his Ph.D. in 2005. He published some 40 papers and was the editor of 2 books. His most important recent publications are: B. Jacobfeuerborn: Reflections on Data, Information and Knowledge. Studia Informatica 2013, 34 (2A), 7–21; B. Jacobfeuerborn, M. Muraszkiewicz: Media, Information Overload, and Information Science. In: R. Bembenik et. al.(eds.) Intelligent Tools for Building Scientific Information Platform. Advanced Architectures and Solutions Berlin: Heidelberg:, Springer 2013, 3–13 [Studies in Computational Intelligence No. 467]; B. Jacobfeuerborn: An Attempt to Innovate Innovation. In. B. Jacobfeuerborn (ed.) Innovating Innovation. Essays on the Intersection of Information Science and Innovation. Warsaw: MOST Press, IINISB University of Warsaw, 2013, 21–40.