Big Data

DOI 10.1007/s12599-014-0338-0

The Authors

Professor Vasant Dhar
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012
USA
vdhar@stern.nyu.edu

Professor Dr. Matthias Jarke (✉)
Information Systems
RWTH Aachen University
Ahornstraße 55
52074 Aachen
Germany
and
Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin
Germany
jarke@dbis.rwth-aachen.de

Dr. Jürgen Laartz
McKinsey & Company Inc.
Budapester Str. 46
10787 Berlin
Germany
juergen.laartz@mckinsey.com

Published online: 2014-08-02

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Vasant Dhar, Matthias Jarke, Jürgen Laartz (2014) Big Data. WIRTSCHAFTSINFORMATIK. doi:10.1007/s11576-014-0428-0.

Physicists and astronomers have been dealing with massive swaths of data for years in order to answer basic questions about the origin and state of the universe. In the late 1990s, Jim Gray initiated an ambitious astronomy data project at Microsoft, identifying issues unique to the new age of "data science" (Microsoft 2009). What does it mean? Is there something unique about it? What skills do *data scientists* need to be productive in a world deluged by data? What are the implications for scientific inquiry?

Data Science is the systematic study of the extraction of knowledge from data (Dhar 2013). Data science aims to discover and extract *actionable* knowledge from the data, that is, knowledge that can be used to make decisions and predictions, not just to *explain* what's going on. The central issues in Data Science have to do with how to endow machines with capabilities that enable them or humans to ask the right questions pertinent to problems based on the flood of heterogeneous data emanating from sensors, humans, and devices. Big data is the grist for Data Science.

The areas that are likely to see impact in the next few decades are social sciences, business, healthcare, entertainment, and sports. Never before have we been as socially connected and tracked as we are now, and never before have we conducted commerce as we do now, with every act recorded indelibly forever. Never before have we had devices that can observe and analyze our every move. This digital exhaust is growing exponentially. Our lives have changed forever, with big data as the engine of discovery, decision making, policy, and the destroyer of privacy as we have known it for centuries.

To put things in historical context, in the 90s, the field of Business Information Systems saw the analysis of historical data collected in data warehouses for reporting and mining (Jarke et al. 2003). In Finance, the availability of increasingly granular data led to automated high-frequency decision making by computers and introduced real-time analysis of streams to data mining. Search engines, web commerce, and social media increasingly added text mining, social network analysis, and heterogeneous data analysis to the mix. In effect, the digitization of business processes in the past two decades has produced exploding data Volumes of much greater Variety with rapidly accelerating Velocity. Beyond these quantitative and technical aspects of the "3V's," Predictive Analytics has emerged as a basis for better human decision making and automated decision making by computers.

Despite the fact that "Big Data" as a buzzword is approaching its peak in the Gartner lifecycle (Dhar 2013), there are reasons to believe that this trend will only accelerate in the coming years. The Internet of Things is likely to lead to further data generation by several additional orders of magnitude through integrating billions of sensors that record images, audio, and video, and ever faster multicore and parallel computers producing ever more complex simulation results that need to be administered and analyzed.

Today, the ubiquity and proliferation of data and analytics is profoundly altering the business landscape. A number of articles have noted that the demand for data scientists is racing ahead of supply (Dhar 2013; Davenport and Patil 2012). The power of analytics is rising while costs are falling. Data visualization, wireless communications and cloud infrastructure are extending the power and reach of information, creating a new foundation for competitiveness. With abundant data from multiple touch points and new analytics tools companies are getting better and better in creating transparency, enabling experimentation to discover actionable insights, exposing variability and improving performance, segmenting populations to customize actions, and innovating/customizing products and services. Companies are learning to test and experiment with big data. They are borrowing from the pioneering efforts of companies such as Amazon and Google, commonly using A/B testing to improve their interfaces and customer experiences. Many are assembling data from real-time monitoring of blogs, news reports, and tweets to detect subtle shifts in sentiment that can affect product and pricing strategy.

Advanced analytics software, increasingly available as a service in the cloud, is making access cheaper and more powerful, enabling machines to identify patterns hidden

in massive data flows or documents. And as companies collect more data from operations, some are looking to create new information-based products that add new revenue streams. In summary big data and analytics is about effecting transformational change in the way a company does its day to day business. In moving towards more evidence based decision making and processes, organizations need to change from a lean continuous improvement mindset to an experimentation driven learning organization. Despite the widespread recognition of big data's potential, organizational and technological complexities as well as the desire for perfection often slow progress.

This special focus issue takes an interdisciplinary business perspective on the Big Data theme. Many of the overall challenges are exposed in the research note "Big Data—An Interdisciplinary Opportunity for Business Information Systems". The paper is based on a broad interdisciplinary survey study conducted by leading German representatives of three the related disciplines business IS, (Krcmar and Schermann, TU Munich), database management (Markl and Hernsen, TU Berlin), and law (Hoeren, Bitter, and Buchmüller, University of Münster) for the German Ministry of Research in 2013, and may be of special interest to many readers because of its clear exposition of Big Data issues in the European legal context.

Three papers were accepted for the special focus issue after several rounds of review and revision, each relating Big Data to major management issues, namely the impact on decision-making styles, the impact on uncertainty handling, and the impact on labor skills requirements.

The first paper, "Big Data and Information Processing in Organizational Decision Processes: A Multiple Case Study" by Kowalczyk and Buxmann (TU Darmstadt, Germany), studies the relationship between the context in which organizational decisions are made, the user of Big Data, and the information processing styles of the decision making. The paper relates the new topic of Big Data to some of the earliest research in organizational information processing. As pure data-centric method application of data analytics might actually increase equivocality in decision making, context-specific combinations of data science methods and organizational decision process elements are proposed.

The second paper, "Taming Uncertainty in Big Data—Evidence from Social Media in Urban Areas" by Bendler, Wagner, Brandt, and Neumann (University of Freiburg, Germany), addresses methods dealing with the inherent uncertainty caused by the much broader range of data sources considered by modern Big Data applications. They use Twitter messages in the urban region of San Francisco to highlight the usefulness of machine-learned spatio-temporal patterns for large numbers of users. Their key idea is the automated and dynamic creation and validation of causal relationships between categories of points of interest, and associated message patterns.

The lack of qualified data scientists may be one of the greatest impediments to progress in the field. What are the required skills, and how do they differ from traditional business intelligence skills? These questions are addressed by an analysis of job postings in the third paper, "Comparing Business Intelligence and Big Data Skills—A Text Mining Study Using Job Advertisements" by Debortoli, Müller, and vom Brocke (University of Liechtenstein). One key observation is that Big Data projects up to now tend to be more research- and human capital intensive than the already established Business Intelligence activities that can rely on well-established standard products and processes. In addition, the world of big data has tended to favor open-source tools compared to the earlier era of business intelligence where the tools tended to be proprietary ones from the established players in IT.

Two interviews with leading researchers/entrepreneurs in the field complement these research papers. Physics professor and entrepreneur Michael Feindt complements the debate on uncertainty management and decision styles by highlighting application domains in which completely automated individualized decision making in digital business outperforms human decision making by what he calls Prescriptive Analytics, whereas research manager Stefan Wrobel adds to the debate on interdisciplinarity and skills management by the example of the Fraunhofer Big Data Alliance, the currently largest applied research organization in this field in Europe.

Acknowledgements

We should like to thank all the authors of submitted papers and the reviewers for their contributions, and the managing editor and editorial office for their guidance and assistance in the preparation process.

References

Davenport TH, Patil DJ (2012) Data scientist: the sexiest job of the 21st century. Harv Bus Rev 90(10):70–76
Dhar V (2013) Data science and prediction. Commun ACM 56(12):64–73
Jarke M, Lenzerini M, Vassiliou Y, Vassiliadis P (2003) Fundamentals of data warehouses, 2nd edn. Springer, Heidelberg
Microsoft (ed) (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond