Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2014 Proceedings

THE IMPACT OF BIG DATA ON THE EPISTEMOLOGICAL DICOURSE IN INFORMATION SYSTEMS RESEARCH

Sebastian Olbrich
University of Duisburg-Essen, Duisburg, Germany, sebastian.olbrich@uni-due.de

Paul Alpar
Philipps University at Marburg, Marburg, Germany, alpar@wiwi.uni-marburg.de

Gad Ariav Recanati
Tel Aviv University, Tel Aviv , Israel, gadia@tauex.tau.ac.il

Opher Etzion
Yezreel Valley Academic College, Israel, opher.etzion@gmail.com

Monica Garfield
Bentley University, Waltham, MA, USA, mgarfield@bentley.edu

THE IMPACT OF BIG DATA ON THE EPISTEMOLOGICAL DICOURSE IN INFORMATION SYSTEMS RESEARCH

Panel

Sebastian Olbrich, University of Duisburg-Essen, Mercator School of Management, Duisburg, Germany, sebastian.olbrich@uni-due.de Facilitator:

Paul Alpar, Philipps University at Marburg, School of Business and Economics, Marburg, Germany, alpar@staff.uni-marburg.de

Gad Ariav Recanati Business School, Tel Aviv University gadia@tauex.tau.ac.il

Opher Etzion, Yezreel Valley Academic College, Israel, opher.etzion@gmail.com

Monica J. Garfield, Bentley University, Waltham, MA, USA, mgarfield@bentley.edu

Abstract

The term Big Data is in the centre of many recent discussions in academia and practice. To date, the discussion is, to a large extent, driven by technological concepts and capabilities; however, a holistic understanding of the term is still missing. This panel intends to investigate the consequences and enabling prerequisites of Big Data. Next to economic and social impacts, the panel addresses questions that will raise issues that impact the academic discourse of Big Data itself.

Keywords: Big Data, Scientific method, hypotheses construction.

"Any piece of knowledge I acquire today has a value at this moment exactly proportional to my skill to deal with it. Tomorrow, when I know more, I recall that piece of knowledge and use it better."

Mark Van Doren (1894–1972)

1 Introduction and motivation

In order to gain competitive advantage in a volatile business environment, large organizations increasingly face the need to maintain and analyze large amounts of structured and unstructured data (Davenport et al. 2007; Laney 2001). The sources of these data can lie inside or outside the organizational borders. The need to access the data may occur on an ad–hoc basis or arise regularly to solve reoccurring decision problems. Neither the trend towards ever more data integration nor the promise of finding value in the amounts of accessible data is really new. Yet, recent developments in social behavior (e.g., heavy use of social media sites), technological progress (e.g., increasing amounts of sensor data becoming available) and the capability to store and analyze these data (enabled mostly by innovation in database technologies) lead together to a new phenomenon summarized by the term 'Big Data'. Hence, Big Data usually includes data sets with sizes beyond the capabilities of commonly used software tools and its users to capture, curate, manage, and process the data within a tolerable elapsed time.

As catchy is this might sound, these and similar introductions to the field of Big Data make the term so difficult to work with. First, taken the term literally it addresses a constantly moving target. For instance, the world's technological per–capita capacity to store information has roughly doubled every 40 months since the 1980s (Hilbert et al. 2011). The previously stated 'commonly used software tools'

also catch up quickly in processing ever more data. There also might be something like 'perceived Big Data' since the perception of a data set as being big might vary in different user groups – let's say from the amount of data processed within the Worldwide LHC Computing Grid (WLCG) at the Conseil Européen pour la Recherche Nucléaire (CERN 2013) to the data needed to establish a marketing campaign at a wholesale company. Second, Big Data are not a sudden phenomenon. It can be considered as a logical development built on technological advances (mostly in database technology and visualization) and a large base of knowledge grounded in various fields such as Operations Research (Churchman et al. 1957; Gass et al. 2005; Kirby 2003), Statistics, Management Science—particularly the decision–making process (Simon 1976) and Management Information Systems (Inmon 1996; March et al. 2007; Shim et al. 2002; Turban et al. 1995). Particularly the field of Business Intelligence (Dresner et al. 2002; Luhn 1958) made it possible to structure the underlying strategies and concepts regarding the integration of large data sets in order to prepare them for analysis and visualization (Moss et al. 2003; Olbrich et al. 2011; Turban et al. 2011; Watson 2010; Wixom et al. 2001). As Peter Norvig—Google's director of research—puts it: "*We don't have better algorithms. We just have more data.*" (McAfee et al. 2012b). Last, as Big Data is about to peak the Gartner hype–cycle (Lapkin 2012) the term would probably have not gotten so much attention among non–specialists if it did not reach certain media attention; particularly due to government involvement. Providing spatial and statistical data, the techniques were initially received positively. As a consequence of the financial crisis, authorities enforced a set of regulations demanding ever more data to be provided to them. Recent court cases in the USA have also led organizations to keep large masses of documents, E–Mails and other forms of electronic communication that may be required if they face litigation. However, the issue has become increasingly controversial as governmental efforts to collect massive private data became public (e.g. Bamford 2012; NSA 2011).

2 Controversial Issues

The demands for data–intensive science represent a challenge for diverse scientific communities (Bell et al. 2009). For that reason, critique of the big data paradigm comes in different flavors:

Epistemological critique. In the mainstream critique about trust in statistics, it is usually pointed out that often very strong assumptions are made about mathematical properties that may not at all reflect what is really going on at the level of micro–processes. That point is also raised in the Big Data discussion, e.g. "*A crucial problem is that we do not know much about the underlying empirical micro–processes that lead to the emergence of the typical network characteristics of Big Data*" (Snijders et al. 2012). Yet, in such fields as health or biology, conventional scientific approaches are based on experimentation. The limiting factor of these approaches is the relevant data that can confirm or refute the initial hypothesis. By the capability of processing ever more data, a complementary postulate is accepted in biosciences and spreads to other disciplines: the processing of huge volumes of data by information systems, without prior hypotheses (Jones et al. 2006). In such approaches, it is the formulation of a relevant hypothesis to explain the data that is the limiting factor (Hey et al. 2009). The search logic is reversed and the limits of induction need to be considered ("Glory of Science and Philosophy scandal" (Broad 1914)).

For academia, particularly the Information Systems discipline, this poses a number of issues. For instance, some consider the lack of ex–ante hypotheses as being the "*end of science"* or at least "*end of theory*" (Anderson 2008). Others point out the social–technical dimension of the phenomenon and the opportunities for researchers since Big Data will always require contextualization in their social, economic, and political contexts (Chen et al. 2012; Eaton et al. 2012; Hilbert 2013). Even as companies invest eight– and nine–figure sums to derive insight from information streaming in from suppliers and customers, it is argued that less than 40% of employees have sufficiently mature processes and skills to do so (Shah et al. 2012). To overcome this insight bottleneck, Big Data, no

matter how comprehensive or well analyzed, need to be complemented by a 'big' degree of judgment (Davenport et al. 2012; McAfee et al. 2012a).

The current execution and interpretation of Big Data can be, and has been, criticized in multiple ways. Currently the political impact of being able to gather and analyze various data types about individuals is in the center of the discussion (e.g. Bamford 2012; NSA 2011). The aspect of data ownership however is not limited to government agencies; the impact and consequences of Big Data analytics for the individual are rarely made transparent (Ohm 2012; Schaar 2009; Wigan et al. 2013). In order for Big Data analytics to be a valuable asset for the organization, multiple organizational issues arise: e.g., who will fill the competency gap (Davenport et al. 2012; Jobanputra 2013), to overcome the current obstacles for BI success (Chenoweth et al. 2006; Hwang et al. 2005; Olbrich et al. 2011; Wixom et al. 2001) and turn big data into sustainable value for organizations (Patil 2012; Redman 2008).

There is also a problem with competition on analytics and belief in Big Data: If a firm concentrates on the use of analytics to fine tune the current business model, it might lose time, money, and effort for the more important task of creating a new business model. Since the data stem from existing processes, they can be used to improve and optimize the current processes and business model but rarely for business model innovation. Thus, continuous improvement based on analytics carries the risk of making companies less flexible and less innovative.

Possible further issues to discuss:

- Framing in IS research: New, independent phenomenon vs. established field (e.g. two tracks for BI/KM and Big Data)
- Definition in terms of the data–focused 3Vs (volume, variety and, velocity) is a rather limited and a technology based view. Where is the social part? Where is KM?
- Concerns about ethics and Data Privacy

To shed light on this issue and to draft a way toward a corresponding research agenda, we propose to discuss these issues in a distinct panel. It is also our aim to steer discussion towards a possible alignment in the IS community, as for instance, the topic of Big Data is subject in several tracks at ECIS and even has a track of its own.

3 The panellists positions

Paul is interested in decision support and intelligent problem solving since his doctoral work on method bases for decision support. Later he researched the application of AI in marketing and production and currently works on collaborative use of BI. He raises the issue that in traditional decision support, we try to solve known problems or develop recognized opportunities. In Big Data analytics, as currently promoted, the hope is that we will stumble on solutions of problems or opportunities that we do not know about. But even when we look for a needle in a hay stack, we know we are looking for a needle. How likely are we to find it, when we do not know that we are looking for, or if we search in the wrong hay stack (i.e., do not use appropriate data)? Therefore, we need to search for ways to merge the approaches for best results. Then, we will be able to improve and innovate business models.

Gad will bring in his more than 45 years of experience in the practice, research and teaching of the use of information technologies and systems in organizations. 12 years of experience in the study and teaching of high-tech entrepreneurship as well as international business strategy and business strategies in technology intensive companies in general. In his view Big Data provides a "third loop"

for data-based value creation - in the concentric structure of an operational loop being the inner circle, decisional is the next and the loop of innovation is the one enabled by Big Data.

Opher will provide input from his experience at the IBM Research Laboratory. Previously he has been lead architect of event processing technology in IBM Websphere, and as Senior Manager in IBM Research division, managed a department that has performed one of the pioneering projects that shaped the area of complex event processing. He is also the chair of the event processing technical society. In his point of view Big Data poses a revolution to the field of analytics. As real time analytics become ever more relevant, the future will tend towards self–learning

Monica has been in the field of knowledge management and creativity for the past 20 years. Her most recent work has focused on how knowledge is created and captured in various environments. The world of big data opens new doors to knowledge creation and capture but it may also block the way to truly new knowledge. Will we forever be trying to navigate into the future by looking through our re–view mirror? How important is understanding the "why" behind correlations and the contexts in which they will hold true? How can big data be used for innovation?

Sebastian, who will facilitate the panel, is motivated by observations he has made in his long experience as a Managing Consultant and an Interim–Manager in the field of Business Intelligence. His motivation for organizing this panel is derived from the observation of an increasing belief in data only. As much as he advocates grounding decisions in sound knowledge bases, he points out the risk of increasing bias of decision makers. In a world of Big Data this dilemma is emphasized by Information Systems which tend to present presumed relationships rather than deliver objective decision support (correlation vs. causality). In his believe this observation can be made in research and in the field.

4 Structure of the panel

The ability to promote an idea to a broad audience is one of the key strength of a panel. Given the rising interest in shaping the society by technical design, we believe that a panel is ideally suited to allow aspiring IS researchers to tap into this debate and better understand some of the controversial methodological issues.

We intend for the panel to have four distinct phases. The major one relies on the participation of the audience. The guiding principles for this procedure are to:

(1) Draw on the panellist's expertise and experience to provide some initial principles and issues on the topic, and

(2) Fully engage participants in advancing these.

For the first phase, the facilitators will present the topic with a few slides featuring a few key figures and an introducing example. Next, the panellists are introduced and their position in the discussion framed. The first phase will take 10–15 minutes (including introductions)

In the second phase, the panellists will have the opportunity to further elaborate on their position and the role in the discussion. Their argumentation will be based on their previous work and experience in the field and on the topic. Each panellist will have between two and three minutes to elaborate her/his position. As such, this phase will take about 8–12 minutes.

The third phase of the panel time will be done in concert with the audience and is organized as an open, yet facilitated, discussion. Since this is the main part of the panel, we plan it to last for about 50–60 minutes. In this phase, the goal is to begin identifying common positions and to identify areas of distinct differences regarding the phenomenon and current issues. In contrast to discussing the concept

of Big Data on an abstract level, we intend to focus the discussion on our discipline and the consequences we draw to shape the future of our discipline.

The fourth phase will last about 10 minutes and will summarize the results and key positions of the discussion. Following the panel, a full summary of the discussion, responses, and points will be posted to either the conference website (if permitted) or to a facilitator's website so that participants can have access to the material they helped to co–create.

5 The Participants' (short) Biographies

Paul Alpar is Professor of Information Systems and Quantitative Methods at the Phillips University at Marburg. He has also been on the faculty of the Univ. of Illinois, Chicago and visiting scholar at Univ. of California, Berkely, Tel–Aviv University, Tel–Aviv, Goethe Univ., Frankfurt/M, Georgia State Univ., Atlanta. He is interested in decision support and intelligent problem solving in various contexts. His research appeared in journals like J. of Management Information Systems, Decision Support Systems, International J. of Research in Marketing, Annals of Operations Research, IEEE Transactions on Engineering Management, International J. of Electronic Commerce, J. of Productivity Analysis. He published about 60 papers in referred journals or proceedings and is the author or co–author of five books. He currently serves as an AE for EJIS and a SE for Information Systems Management.

Gad Ariav looks back on almost 30 years of international academic career, primarily at Recanati Faculty of Management (Tel Aviv University), but also Stern School of Business (NYU), Paul Merage School of Business (UC Irvine), Claremont Graduate University, as well as Rotterdam School of Management (Erasmus University). Recently he spent time as Resident Scholar at the Indian Institute of Management, Bangalore.Served as the academic director of Recanati's executive MBA programs between 1994 and 2005, founding director of the Kellogg-Recanati International EMBA program as well as the Georges Leven High-Tech Management School (*HTMS*), all at Tel Aviv University. His current academic focus lies on the study of global management of innovation and the use of IT for decision support. He published among others in the ACM Transactions on Database Systems, Communications of the ACM, Decision Support Systems, Journal of MIS, as well as the proceedings of the International Conference on Information Systems.

Opher Etzion Opher Etzion serves as Professor of IS, and head of the Technological Empowerment Institute (under construction) in YVC. During the years 1997–2014 he served in various roles in IBM, most recently IBM Senior Technical Staff Member and chief scientist of event processing in IBM Haifa Research Lab,previously he lead architect of event processing technology in IBM Websphere, and a Senior Manager in IBM Research division, managed a department that has performed one of the pioneering projects that shaped the area of "event processing". He has also been the chair of EPTS (Event Processing Technical Society). In parallel he serves as an adjunct professor at the Technion – Israel Institute of Technology, where over the years he supervised 6 PhD and 22 MSc theses. He has authored or co–authored more than 80 papers in refereed journals and conferences, on topics related to: active databases, temporal databases, rule–base systems, event processing and autonomic computing, and gave several keynote addresses and tutorials. He is the co–author of Event Processing in Action (with Peter Niblett), a comprehensive technical book about event processing

Prior to joining IBM in 1997, he has been a faculty member and Founding Head of the Information Systems Engineering department at the Technion, and held professional and managerial positions in industry and in the Israel Air–Force. He is a senior member of ACM, and has been general chair and program chair of various conferences such as COOPIS 2000 and ACM DEBS 2011. He won several prestigious awards over the years, such as the Israel Air–Force commander award, the highest air–force award , IBM Outstanding Innovation Award (twice) , and IBM Corporate Award (the highest

IBM award) for the pioneering work on event processing. He was recognized as Distinguished Speaker by ACM.

Monica Garfield is an Associate Professor in Computer Information System at Bentley University (USA). She holds a PhD from the University of Georgia, and MS and MBA from Boston University and an AB from Vassar College. Her research interests include the use of IT for group collaboration toward the goal of knowledge creation and the fostering of creativity. She teaches database management systems at the graduate and undergraduate level and strategic information systems at the graduate level. Her work has been published in MIS Quarterly, ISR, JMIS, and Journal of Information and Knowledge Management,

Sebastian Olbrich who will facilitate the panel, is a post–doc researcher in the field of Business Analytics and Corporate Performance Management at the Mercator School of Management (MSM), University of Duisburg–Essen. He also has a long experience as a Managing Consultant and an Interim–Manager in the field of Business Intelligence. Next to his work as a practitioner he published 50+ Academic articles, serves for IS conference and Journals in various roles and is speaker at various BI and CPM related conferences such as TDWI, Management Circle, European Finance Week, IBM Business Connect or CeBIT.

References

Anderson, C. "The End of Theory The Data Deluge Makes the Scientific Method Obsolete," *Wired Magazine* (16.07) 2008.

Bamford, J. "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)," *WIRED* (2013:03) 2012.

Bell, G., Hey, T., and Szalay, A. "Beyond the Data Deluge," Science (323:5919) 2009, pp 1297–1298.

Broad, C.D. Perception, physics and reality. An Enquiry into the Information that Physical Science can Supply about the Real Cambridge University Press, London, 1914.

CERN "Welcome to the Worldwide LHC Computing Grid ", WLCG Office, Geneva, Switzerland, 2013.

Chen, H., Chiang, R., and Storey, V. "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly* (36:4) 2012, pp 1165–1188.

Chenoweth, T., Corral, K., and Demirkan, H. "Seven Key Interventions for Data Warehouse Success," *Communications of the ACM* (49:1) 2006, pp 114–119.

Churchman, C.W., Ackoff, R., and Arnoff, L. *Introduction to Operations Research* Wiley, 1957.

Davenport, T.H., and Harris, J.G. *Competing on Analytics: The New Science of Winning* Mcgraw–Hill Professional, Massachusetts, 2007.

Davenport, T.H., and Patil, D.J. "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review* (Oktober 2012) 2012.

Dresner, H.J., Buytendijk, F., Linden, A., Friedman, T., Strange, K.H., Knox, M., and Camm, M. *The Business Intelligence Competency Center: An Essential Business Strategy* Gartner Research, Stamford, 2002.

Eaton, C., DeRoos, D., Deutsch, T., Lapis, G., and Zikopoulos, P. "Understanding Big Data ", McGraw Hill, New York, 2012.

Gass, S.I., and Assad, A.A. *An Annotated Timeline of Operations Research: An Informal History* Kluwer Academic Publishers, New York, 2005.

Hey, T., Tansley, S., and Tolle, K. *The Fourth Paradigm– Data–Intensive Scientific Discovery* Microsoft Research, Remond, CA, 2009.

Hilbert, M. "Big Data for Development: From Information–to Knowledge Societies ", Social Science Research Network Rochester, NY.

Hilbert, M., and López, P. "The World's Technological Capacity to Store, Communicate, and Compute Information," *Science* (332:6025) 2011, pp 60–65.

Hwang, M.I., and Hongjiang, X. "A Survey of Data Warehousing Success Issues," *Business Intelligence Journal* (10:4) 2005, pp 7–14.

Inmon, W.H. *Building the data warehouse*, (2 ed.) Wiley, New York, NY, 1996, p. 401.

Jobanputra, D. "Editorial: Taking the lead," *The Actuary* (03:October) 2013, p 5.

Jones, M., Schildhauer, M., Reichman, O., and Bowers, S. "The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere," *Annual Review of Ecology, Evolution, and Systematics* (37:1) 2006, pp 519–544.

Kirby, M.W. Operational Research in War and Peace: The British Experience from the 1930s to 1970 Imperial College Press, 2003.

Laney, D. "3D Data Management: Controlling Data Volume, Velocity and Variety," META Group Report, 2001.

Lapkin, A. "Hype Cycle for Big Data, 2012 ", Gartner Group.

Luhn, H.P. "A Business Intelligence System," *IBM Journal of Reserach an Development* (2:4) 1958.

March, S.T., and Hevner, A.R. "Integrated Decision Support Systems: A Data Warehousing Perspective," *Decision Support Systems* (43:3) 2007, pp 1031–1043.

McAfee, A., and Brynjolfsson, E. "Big Data: The Management Revolution," *Harvard Business Review* (Oktober) 2012a.

McAfee, A., and Erik Brynjolfsson, E. "Big Data: The Management Revolution," *Harvard Business Review* (Oktober) 2012b.
Moss, T.L., and Atre, S. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision–support Applications Addison–Wesley, Boston, 2003.
NSA "Groundbreaking Ceremony Held for $1.2 Billion Utah Data Center," NSA Press Release (6 January 2011), 2011.
Ohm, P. "Don't Build a Database of Ruin," Harvard Business Review, 2012.
Olbrich, S., Niehaves, B., and Pöppelbuß, J. "BI Systems Managers' Perception of Critical Contextual Success Factors: A Delphi Study," International Conference on Information Systems (ICIS), Shanghai, China, 2011.
Patil, D.J. (ed.) Data Jujitsu: The Art of Turning Data into Product. O'Reilly Media, 2012.
Redman, T.C. Data Driven: Profiting from Your Most Important Business Asset Harvard Business Review Press 2008.
Schaar, P. "25 Jahre Volkszählungsurteil. Datenschutz – Durchstarten in die Zukunft!," Der Bundesbeauftragte für den Datenschutz und die Informationssicherheit, Bonn, 2009.
Shah, S., Horne, A., and Capellá, J. "Good Data Won't Guarantee Good Decisions," *Harvard Business Review* (8 September) 2012.
Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., and Carlsson, C. "Past, Present, and Future of Decision Support Technology," *Decision Support Systems* (33:2) 2002, pp 111–126.
Simon, H. *Administrative Behavior*, (3rd ed.) The Free Press, New York, 1976.
Snijders, C., Matzat, U., and Reips, U.–D. "'Big Data': Big gaps of knowledge in the field of Internet," *International Journal of Internet Science* (7) 2012, pp 1–5.
Turban, E., Sharda, R., Delen, D., and Aronson, J.E. *Decision support and business intelligence systems*, (9 ed.) Pearson, Upper Saddle River, NJ, 2011, pp. XXII, 696 S.
Turban, E., and Walls, J.G. "Executive information systems—a special issue," *Decision Support Systems* (14) 1995, pp 85–88.
Watson, H.J. "BI–based Organizations," *Business Intelligence Journal* (15:2) 2010.
Wigan, M.R., and Clarke, R. "Big Data's Big Unintended Consequences," *Computer* (46:6) 2013, pp 46–53.
Wixom, B.H., and Watson, H.J. "An Empirical Investigation of the Factors Affecting Data Warehousing Success," *MIS Quarterly* (25:1) 2001, pp 17–41.