Big Data: A fashionable topic with(out) sustainable relevance for research and practice?

by

Hans Ulrich Buhl, Florian Moser, Julia Heidemann, Maximilian Röglinger

March 2013

Big Data

A Fashionable Topic with(out) Sustainable Relevance for Research and Practice?

The Authors

Prof. Dr. Hans Ulrich Buhl (✉)
Dr. Maximilian Röglinger
Dipl.-Kfm. Florian Moser
FIM Research Center Finance &
Information Management
University of Augsburg
Universitätsstraße 12
86159 Augsburg
Germany
hans-ulrich.buhl@wiwi.uni-augsburg.
de
maximilian.roeglinger@wiwi.uni-
augsburg.de
florian.moser@wiwi.uni-augsburg.de

Dr. Julia Heidemann
McKinsey & Company, Inc.
Sophienstr. 26
80333 Munich
Germany
julia_heidemann@mckinsey.com

"*To make money, you've got to predict two things—what's going to happen and what people think is going to happen.*"

When looking at the words of Hal Varian, Google's Chief Economist and professor emeritus at the University of California, Berkeley, thinking of Big Data seems natural. Big Data – a dictum which currently seems to be on everyone's lips – has recently developed into one of the most discussed topics in research and practice. Looking at academic publications, we find that more than 70 % of all ranked papers which deal with Big Data were published within the last two years (Pospiech and Felden 2012) as well as nearly 12,000 hits for Big Data on GoogleScholar across various fields of research. In 2011, more than 530 academic Big Data related publications could be counted (Chen et al. 2012). We find more hits for "Big Data" than for "Development aid" in Google, and almost daily an IT-related business magazine publishes a Big Data special issue next to a myriad of Big Data business conferences. In Gartner's current Hype Cycle for Emerging Technologies (Gartner 2012), Big Data is right on the peak of its hype phase, and according to this source a broad adoption is to be expected within the next five years. Big Data provokes excitement across various fields such as science, governments, and industries like media and telecommunications, health care engineering, or finance where organizations are facing a massive quantity of data and new technologies to store, process, and analyze those data. Despite the cherished expectations and hopes, the question is why we face such excitement around Big Data which at first view rather seems to be a fashionable hype than a revolutionary concept. Is Big Data really something new or is it just new wine in old bottles seeing that, e.g., data analytics is doing the same type of analysis since decades? Do more data, increased or faster analytics always imply better decisions, products, or services, or is Big Data just another buzzword to stimulate the IT providers' sales? Taking the traditional financial service industry, which currently cherishes huge expectations in Big Data, as an example, the collection of massive amounts of data via multiple channels for a long time was part of the business model to customize prices, product offers, or to calculate credit ratings. However, improving financial services by exploiting these huge amounts of data implied constant updating efforts, media disruptions and expensive acquisition and processing of data. Hence, more data resulted in expensive data management, in higher prices for products or services as well as in inconvenient processes regarding the customers' data entry. Hence, instead of the traditional universal banks that focused on a data-intensive business model, direct banks with a higher grade of standardization and IT support as well as a focus on (very few) key customer data often enough have become more successful. Focusing solely on pure IT-based data acquisition, processing and analysis to save costs on the other side is virtually impossible in industries such as banking due to an intense personal contact. Besides, neither in the financial service industry nor in other industries do more data automatically lead to better data, better business success, better services, better decisions, or (more) satisfied customers. Above all, Big Data brings a lot of still unresolved challenges regarding the volume, velocity, variety, and veracity of data, which should not be underestimated. Often enough, more data even lead to a certain amount of "data garbage" which usually is more easily and better recognized and managed by employees rather than by analytics software (veracity). Additionally, the management of various sources of data such as from, e.g., mobile applications, online social networks, or CRM systems is far from trivial (variety). The high data traffic brings along the challenge of archiving, retrieving, and analyzing huge amounts of data in real-time (volume and velocity). Unsurprisingly, nearly every second Big Data project is canceled before completion (Infochimps 2013). And as if these challenges were not enough, we additionally see a myriad of different legal privacy restrictions in different countries turning into one of Big Data's most serious challenges.

Despite a customer generation which is increasingly losing its inhibition to distribute private data in virtually every place in the web, country-specific privacy laws and a significant number of customers who are not willing to have their private data stored for a long time might seriously impede Big Data approaches and threaten corresponding business models.

In view of this development, is Big Data really "the next big thing" with substantial economic impact and technological significance within the next decade as currently promoted in research and practice? Yes and No – although Big Data currently might provoke exaggerated expectations, labeling it purely as a fashionable topic for existing concepts may be just the easy way out when we consider the following developments: The amount of data that is produced each day already exceeds 2.5 exabytes (McAfee and Brynjolfsson 2012). The bidirectional telecommunications capacity is growing by almost 30 % per year and the globally stored information is increasing more than about 20 % per year (Hilbert and López 2011). The turnover in the Big Data segment is expected to increase by more than 400 % to 16 billion Euros in 2016 (Computerwoche 2012). Looking at these points, the relevance of Big Data for academic and practical initiatives which regard this new era of data as a chance rather than a threat becomes obvious. As in his view data are becoming one of tomorrow's most valuable goods for offering adequate products or services to customers, Jürgen Fitschen (Co-Chairman Deutsche Bank) even considers companies like Google or Microsoft as Deutsche Bank's main future competitors of tomorrow (Deutsche Bank 2013). Indeed, driven by technological developments like, e.g., mobile and sensor-based content, various possibilities emerge for companies and governments (e.g., market intelligence, public safety) as well as for research (e.g., network analytics, mobile analytics) (Chen et al. 2012). In fact, some companies such as the mail order company Otto already exploit their huge volume of data successfully. On the basis of more than 300 million data sets per week, Otto performs more than one billion forecasts per year to predict the sales of certain articles in the next days and weeks. This enables Otto to decrease the inventories on average by 30 % (Fischermann and Götz 2013). Others such as the American broadband and telecommunications company Verizon have more visionary ideas which almost go into the direction of an Orwellian society. Verizon has applied for a patent in which a home entertainment system sends couple therapy advertisements to a television or a mobile device as soon as it recognizes a couple arguing. Alternatively, the system sends advertisements for a romantic weekend or contraceptives in case the couple is cuddling (Fischermann and Götz 2013). Of course, future data's volume, velocity, variety, and veracity as well as privacy concerns might appear as stumbling blocks for such wishful thinking of Big Data. However, in consideration of the following technological developments and internal efforts regarding data quality and privacy issues, companies might be able to pave the way for their individual Big Data success:

1. Big Data is driven by massive cost reduction in data management in combination with Moore's law regarding processing power. New technologies such as, e.g., Quantum Computing or In Memory Database systems allow for handling new dimensions of data amounts quickly and in an economically efficient way (volume and velocity). However, it is critical to align new IT infrastructure opportunities with existing and new business processes and applications in order to be able to exploit technological infrastructure advancements.
2. Successful Big Data approaches require new tools such as e.g., Social, In-Memory, Text, or Semantic Analytics which allow for analyzing the new amount of different data sources from for instance online social networks, search engines, payment transactions, or all kinds of E-Commerce (variety). However, the application of such data analytics tools first requires the possibility to gain access to these new data and customer sources as well as their adaption the new data sources to existing data warehouses, reporting standards etc.
3. Big Data's success is inevitably linked to an intelligent management of data selection and usage as well as joint efforts towards clear rules regarding data quality. Though new technologies allow for collecting more and more data, the future customer is not likely to be willing to enter various kind of data, e.g., in mobile product purchase. Future applications need to hold 99 % of customer data always available from various sources and only 1 % to be entered on demand by the customer. This requires high quality of the data held by the company to guarantee meaningful use of the new data entered by the customer. High data quality requires data to be consistent regarding time (e.g., across all sales channels), content (e.g., same units of measure), meaning (e.g., to avoid different meanings), and data that allow for unique

identifiability (e.g., of customers), as well as being complete, comprehensible, and reliable. To this end, a clear data governance and data policy is inevitable which enables a meaningful use of the data (veracity). As data policies likely differ e.g., within different business units or countries, companies need a data governance with clear data quality policies, data quality management processes, data quality responsibilities etc. In absence of this condition, all technological infrastructure advancements, analytic tools or business models are ultimately without value for data-driven business decisions.

4. Big Data requires innovative approaches which view privacy concerns and different international privacy standards not as hindering restrictions, but rather as a chance to develop a competitive advantage. In a Big Data era with many different data from different sources, privacy and anonymity means more than just uncoupling surname, first name, age, and address from a dataset. Location-based data and other sources still allow for easy and clear identification and tracking.

With respect to privacy, we can still observe (too) many companies especially from Europe and Asia avoid making first moves in Big Data. Rather than to wait for the well-known global companies like Google, Amazon, or Facebook to make the first step, it is time for small and medium-sized companies worldwide to become leaders in this new emerging business area. Otherwise, we will see a second wave of digital "colonialization" and domination by these Internet giants, as we have seen after the dot-com bubble where many companies worldwide were afraid of investing in new risky business models. Certainly, companies such as Google or Facebook are in the happy position of not having to deal with the restrictions of strict privacy policies in their domestic markets compared to the manifold legal restrictions which, e.g., German companies face. Thus, at a first glance they might be ahead regarding the usage of data. However, restrictions in certain markets not always need to be disadvantageous for long-term success of an industry. Looking for instance back at the development in the automotive industry, German manufacturers far earlier had to deal with customers expecting both fuel-saving cars and high performance driving experience whereas US-based manufacturers did not have to care about fuel-efficiency due to low fuel prices in their domestic market. Today, all customers worldwide face rising fuel prices and are developing a stronger ecological awareness. The German manufacturers' know-how in building fuel-efficient cars today is one of the reasons why Germany's automotive industry outperforms its American counterpart. Whereas the US-based companies struggle with low market shares and a bad image in the last decades, the German manufacturers dominated the global markets and regularly realize sales records (above all in growing markets and the US). Hence, constraints can actually serve as a fertile stimulus for innovative, customer oriented and value creating solutions. Regarding Big Data, the restrictive privacy rules, e.g., in Germany can be a similar chance for companies to develop innovative business models that satisfy legal privacy restrictions and customers' concerns and simultaneously create value for the company.

Thus, to benefit from Big Data requires changes and improvements of technological infrastructure, business processes, business applications as well as an incremental change in the business model of the company, including new methods to derive knowledge from data. Companies aiming at a better use of gathered data should regard this also as a cultural challenge and, e.g., focus on training employees to efficiently manage data properly and incorporate them into decision making processes. Rather than considering data simply as an input variable, their value as a company "asset" needs to be understood. To maximize this goods' utility, data governance must ensure high data quality as required as a basis for any Big Data initiative. This might also imply the creation of some new roles, such as a Chief Data Officers or data scientists as well as rules for data-driven decisions. In view of the fact that the majority of companies rates its level of data integration maturity as (very) low or average (Forrester Research 2010), there is still a long way to go before companies can utilize Big Data in the way currently promoted.

As a consequence, practice is well advised to closely collaborate with researchers who follow a multidisciplinary research approach to deal with Big Data's variety of challenges and to derive lasting business models. This inevitably raises the question of the role of Business and Information Systems Engineering (BISE) in the Big Data debate and its possible contribution to making Big Data a success without just jumping on the hype bandwagon. The design of effective and efficient application systems to process and manage large amounts of data in fact is nothing new for BISE as its former

names "elektronische datenverarbeitung" (Electronic Data Processing) or "Angewandte Informatik" (Applied Computer Science) reveal. Also the data-oriented research topics during the 1970s and 1980s, shortly after companies like Software AG (1962) or SAP (1972) were founded, show a solid data-oriented research agenda that now needs to be revived and adjusted. However, new ways in analytics will in the nearer future also require a close collaboration with Operations Research, one of BISE's most important neighboring disciplines. On our own behalf: The re-staffing of BISE's editorial board goes in line with this future trend as it incorporates a broader range of researchers from different fields and so allows for a closer relationship with Operations Research and Operations Management research. Taking BISE's variety of methods and its nature as a multidisciplinary research discipline as a basis, BISE can contribute to the development of Big Data from both a theoretic and a practical perspective in two domains:

1. New research opportunities: The emerging challenges of Big Data's volume, velocity, variety and veracity as well as quality and privacy issues pose various research questions regarding the implications for data storage, processing and analysis, changes in applications and processes as well as new business model opportunities. Apart from very technological issues like database optimization or semantic analysis, also the management and the optimization of business processes, the economic valuation of Big Data business models and many more (e.g., privacy-preserving data analytics) need to be addressed. To this aim, BISE needs to build on its broad variety of research methods to foster its claim of being the leading research community which offers well-founded theoretical solutions transferable into practice to address the broad range of Big Data challenges.

2. New teaching aspects: In contrast to earlier times when most business decisions relied on internal and transactional data, the business decisions of tomorrow require the involvement of huge volumes of more and more external information and take place outside the IT functions (Chen et al. 2012). This and the variety of challenges, fields of applications and importance in nearly every industry calls for data experts on the one hand, but even more for well-founded and multidisciplinary knowledge of future talents and leaders on the other hand. Solely for the US, a shortfall of 1.5 million managers with know-how for data-driven decision making is expected by 2018 (Manyika et al. 2011). This improves BISE's attractiveness for students from various fields such as, e.g., business, information systems, (business) mathematics, computer science or (industrial) engineering as a discipline to study or do research in. However, though many claim that data scientist is becoming the most demanded and sexiest job in the world, it is up to the BISE community to educate talents which are able to build bridges between theory and practice as well as able to deal with both technical and economic questions.

In conclusion, Big Data – besides all hype and cherished expectations as "the next big thing" – above all is a multidisciplinary and evolutionary fusion of new technologies in combination with new dimensions in data storage and processing (volume and velocity), a new era of data source variety (variety) and the challenge of managing data quality adequately (veracity). However, to render Big Data a worthwhile innovation rather than merely a gadget, companies need well-founded and innovative business models that create value for the customer and thus the company while simultaneously considering privacy restraints. Hence, both from the research and practice perspective, Big Data needs to be taken as the basis rather than a guarantor of success. For long-term success, IT infrastructure, business processes, applications as well as the business model focusing on the customer need to be completely aligned.

References

Chen H, Chiang RHL, Storey VC (2012) Business intelligence and analytics: from big data to big impact. MIS Quarterly 36(4):1165–1188

Computerwoche (2012) Big data – big business. Computerwoche 43

Deutsche Bank (2013) Jürgen Fitschen im Interview mit der Börsenzeitung: "Wir brauchen paneuropäische Banken". https://www.deutsche-bank.de/de/content/company/nachrichten_4949.htm. Accessed 2013-01-04

Fischermann T, Götz H (2013) Wer hebt das Datengold? Die Zeit, No 2

Forrester Research (2010) Global master data management online survey

Gartner (2012) Gartner's 2012 hype cycle special report evaluates the maturity of 1,900 technologies. http://www.gartner.com/technology/research/hype-cycles. Accessed 2012-11-08

Hilbert M, López P (2011) The world's technological capacity to store, communicate, and compute information. Science 332:60–65

Infochimps (2013) Intelligent applications: the big data theme for 2013. http://blog.infochimps.com. Accessed 2013-01-04

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation. Accessed 2013-01-03

McAfee A, Brynjolfsson E (2012) Big data: the management revolution. Harvard Business Review

Pospiech M, Felden C (2012) Big data – a state-of-the-art. In: Proceedings of the 18th Americas conference on information systems (AMCIS), Seattle