Big Data Challenges

Impact, potential responses and research needs

Daniel Bachlechner
Competence Center Emerging Technologies
Fraunhofer Institute for Systems and Innovation Research
Karlsruhe, Germany
daniel.bachlechner@isi.fraunhofer.de

Timo Leimbach
School of Communication and Culture
Aarhus University
Aarhus, Denmark
timo.leimbach@cc.au.dk

***Abstract*—Although reports on big data success stories have been accumulating in the media, most organizations dealing with high-volume, high-velocity and high-variety information assets still face challenges. Only a thorough understanding of these challenges puts organizations into a position in which they can make an informed decision for or against big data, and, if the decision is positive, overcome the challenges smoothly. The combination of a series of interviews with leading experts from enterprises, associations and research institutions, and focused literature reviews allowed not only identifying and describing eight key challenges but also characterizing their impact, outlining potential responses to them and proposing directions for future research. The challenges faced were found to be not only technological in nature. Organizational and people-related matters as well as the legislative framework are also relevant. For large enterprises and startups specialized in big data, it is typically easier to overcome the challenges than it is for other enterprises and public administration bodies.**

Keywords—big data; challenges; interviews; impact; responses; research directions

I. INTRODUCTION

According to Gartner, 'big data' is a term used to refer to high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making [1]. The first part of the definition refers to the three core characteristics of big data coined by Laney [2]. The characteristics, which are also known as the '3 Vs', are volume (scale of data), velocity (analysis of streaming data) and variety (different forms of data). A fourth characteristic of big data considered important today is veracity (uncertainty of data) [3]. The second part of Gartner's definition highlights the relevance of both the relation between costs and outcomes, and new technological capabilities. Finally, the third part of the definition refers to the ultimate goal of creating value through the processing of data. Value is deemed a fifth characteristic relevant in the context of big data [4].

There is a plethora of application fields for big data [5, 6]. With respect to enterprises, studies name fields such as customer targeting or business process improvement [7, 8]. For research institutions, big data has already been relevant for years [9, 10] and for public administration bodies, it is of growing importance [11]. Every individual has access to more data today than anyone would have imagined one or two decades ago. Data is increasingly becoming a key factor of production. Within the scope of the digital economy, which is also referred to as 'data economy' [12], data is even considered a new currency [13, 14].

Its omnipresence and relevance is leading to entirely new business models [15]. Goods are more and more subject to price pressure and thus the offering of services related to them gains importance [16, 17]. Data is considered a prerequisite to provide such services. There is a trend, however, towards the centralization of data at a few organizations [18]. Currently, enterprises such as Google, Akamai and Facebook enjoy near-monopolies with respect to certain services or types of data. The notion of data marketplaces, which aggregate data and make it available for analyses, is considered a promising foundation for new business models in the context of big data [19, 13].

An entirely new market of big data technologies and services has emerged over half a decade to help capture and extract value from data. Both highly specialized startups and big players in the information technology (IT) industry increasingly aim at capturing a share of this market, which is highly dynamic and currently dominated by US enterprises. IDC analysts expect the worldwide market of big data technologies and services to grow at a compound annual growth rate of 23% through 2019, when it reaches a size of $48.6 billion [20]. The revenue from big data technologies and services, however, is small compared to the business and societal value that is expected to result from enterprises across all sectors, public administration bodies and research institutions that increasingly have the tools at their disposal to make innovative use of data.

However, despite a number of success stories that attracted media attention, most organizations dealing with big data still face considerable challenges. We suppose that only a thorough understanding of these challenges allows organizations, first, make an informed decision for or against big data, and, second, if the decision is positive, overcome the challenges smoothly. So far, research on challenges faced by organizations dealing with big data has largely been focused on technological aspects [21–24, 5]. Other aspects, however, have not yet been discussed much in an academic context.

The research leading to the presented results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 645244.

The paper is structured as follows: The data collection and analysis methods applied are explained in section 2. In section 3, the results are presented putting emphasis on the key challenges themselves as well as on their impact and potential responses to them. Section 4 generates an overall picture, discusses the challenges in the light of related literature and outlines directions for future research. The major conclusions are summarized in section 5.

II. METHODS

A series of expert interviews was conducted to gain insight into perceptions of big data challenges. Systematizing expert interviews were selected for data collection as they had already been used successfully to reconstruct expert knowledge in other fields [25]. Among the deliberately chosen experts were two representatives from enterprises, two from associations and three from research institutions. When selecting the experts, we aimed at making sure that different points of view were taken into account. While one of the enterprises was a big player in the IT industry, which entered the big data market a few years ago, the other was a specialized big data startup. The respective interviewees were senior executives. While one of the associations focused on privacy and civil liberties, the other represented industry interests. The respective interviewees were senior member and chair, respectively. Among the interviewees representing research institutions were one head and two senior members of university departments focusing on databases, information systems and IT law, respectively.

At the time of the interviews, all interviewees could look back on several years of experience with big data. The interviews were semi-structured to allow comparing and aggregating the data collected. To break the ice, the interviewees were first asked to briefly explain their background. Subsequently, key big data challenges were collected. Challenges which had been identified in prior research were brought to the discussion table by the interviewer in case they were not mentioned by the interviewee. Then, the impact of the challenges and potential responses to them were discussed. Finally, we gathered research needs in the context of big data. The interviews were held via telephone, lasted between one hour and one hour and a half, and were, with the consent of the interviewees, digitally recorded.

The interviews were transcribed and analyzed qualitatively. We selected directed content analysis as data analysis method [26]. Accordingly, the coding categories were not derived directly from the transcripts but, in addition to the research questions, findings of prior research served as guidance for the development of initial categories. Big data challenges, identified within the scope of quantitative studies, were used as a starting point for the analysis. These studies, however, neither provide profound descriptions of challenges, nor do they shed light on aspects such as impact or responses. While the interview results cannot be viewed as representative for any specific population, a rich description of particularly relevant challenges is what they can provide. All transcripts were reviewed carefully and text, which appeared to describe a challenge, was coded, whenever possible, using the predetermined categories. Relevant text, which could not be coded with one of these categories, was coded with a new category that captured the essence of the challenge. Afterwards, the texts coded into each category were examined to determine whether subcategories were needed as well as whether categories could be merged. Subsequently, the texts coded into specific categories were compared and aggregated. The analysis allowed identifying commonalities and divergences in the views of the interviewed experts.

The series of interviews was complemented by focused literature reviews. First, for the preparation of the interviews and their analysis, prior research on big data challenges was collected and evaluated. Quantitative studies [7, 8, 27–29] were found to be particularly useful in this regard. The studies taken into account, however, mainly focus on enterprises. Second, once the key challenges were identified and described, focused literature reviews were conducted to identify the main directions in current research as well as promising avenues for future research.

III. RESULTS

Based on the expert interviews, which were informed by prior research, it was possible not only to identify challenges associated with big data but also to gain insight into their impact and responses to them. Eight key challenges, to which the interviewees attached particular importance, are introduced in subsection A. In subsection B, the impact of the challenges as well as potential responses to them are outlined.

A. Key Challenges

The sequence in which the challenges are introduced was set to best allow illustrating relationships between the challenges. It does not reflect the amount of support the challenges received in the interviews or in prior research.

'Identifying use cases' was, for instance, deemed a key challenge by the interviewees. Although big data has reached extensive media attention, the current state of knowledge regarding the topic is rather low. One of the interviewed enterprise representatives stated, *"Although we are a technology provider, people often approach us saying they do have data but don't know what to do with it. They want us to identify use cases as well as to construct business models to allow them justify their big data investment."* Accordingly, the information systems researcher explained that people talking about big data usually have, if any, very specific use cases in mind including, for instance, real-time marketing on mobile devices or sentiment analyses of social network data because they are regularly used by technology providers for advertisement purposes. The average organization, however, has other requirements. It typically needs to build a business case showing how big data creates added value with respect to core operational processes. The researcher stressed, *"Simply being able to increase the velocity of an analysis doesn't create added value. To create added value, it is usually necessary to come up with new use cases where data sets are linked, which haven't been linked before."*

Moreover, the interviewees considered 'setting up a technical architecture' particularly challenging. One of the

interviewees representing an enterprise explained that no two use cases result in the same architectural requirements as well as that individual big data solutions are highly differentiated. The well-known framework Hadoop, for instance, was designed for batch processing of high-volume data but not for highly interactive applications. Consequently, an architecture integrating various big data solutions is usually required. The interviewee stated, *"The challenge is to find the right mix of big data solutions. It is very unlikely that all requirements will be met by a single solution."* Setting up a technical architecture proved to be very costly and complex, particularly, if an entire chain of use cases is to be supported. However, it gets even more challenging as soon as an architecture has to be adapted to new use cases or changing architectural requirements. The expert further pointed out, "*Without doubt, there is a matching solution for every use case, it 'just' has to be identified and integrated into the architecture. […] It becomes really difficult, however, if solutions are used in a way which they were not designed for."* According to the interviewee, this could lead to serious performance issues.

'Carrying out advanced analytics' was considered another key challenge. The big data solutions available today do not support concepts such as 'iteration', which is required, however, to carry out specific clustering, classification or network analysis tasks. Relational databases, for instance, do not allow the repeated execution of a block of statements until some termination criteria are met. The databases researcher stated, *"If we go beyond simple aggregations and transformation of data, if we want to use data mining methods, which are required, for instance, for carrying out predictive or prescriptive analytics […], we reach the limits of current big data technologies."* Hadoop, for instance, does not natively support iterations and, consequently, does not allow executing iterative algorithms efficiently. To execute iterative algorithms at all, drivers have to be programmed manually for each individual case. Additionally, algorithms still have weaknesses with regard to scalability, particularly, if heterogeneous data is processed. The interviewee stressed that algorithms have to be adapted, for instance, not only to graph data but also to audio and video data.

'Promoting cooperation' was also deemed particularly challenging. Interdepartmental and interpersonal cooperation facilitates the identification of data sets whose linkage could be rewarding and, consequently, the identification of use cases that create added value. The information systems researcher stated, *"Big data, particularly if the emphasis in on variety, works only, if something […] is proposed by members of the organization. The idea of demand pull fits much better than the one of technology push."* While linking data from different organizational departments typically imposes high burden on all departments involved as they have to prepare the data according to the relevant requirements, it is not self-evident that all departments benefit from the linkage of the data. One of the enterprise representatives emphasized that big data is still considered a technology topic rather than as a strategic topic with overarching relevance.

The interviewees also considered 'hiring competent experts' a key challenge. One of the interviewees representing an enterprise stated, *"Experts who can operate big data solutions are rare. Even if an organization would want to take advantage of big data […], it is extremely difficult to hire the competent experts needed."* The experts, which are mostly referred to as data scientists, have to be well versed not only with respect to the areas of data analysis and scalable data management but also with respect to the application domain relevant for the organization they work for. Apart from the relative novelty of big data processing, the lack of competent experts can be traced back to the considerable demands placed on them. The concept 'T-shaped professional' is used to describe persons having not only in-depth knowledge in one field but also wide general knowledge. In the big data context, such persons do not only have to be able to set up the technical architecture but also to prepare and execute analyses meeting the needs of the functional managers. The databases researcher emphasized that data scientists and others involved need to be aware of veracity issues in the context of big data. The expert stated, *"The worst thing one can do is to accept the results of big data analyses uncritically. There is always some uncertainty. One needs to question the results applying common sense."* The interviewee used the term 'information literacy' to refer to the ability of persons to critically evaluate data and the results of their analysis.

Moreover, the interviewees deemed 'making solutions consumable' particularly challenging. According to the information systems researcher, the big data solutions available today cannot be used by most of the 'normal' staff in the organizational departments. One of the enterprise representatives stated that even business analysts usually cannot use them or adapt them to new use cases. The databases researcher explained, *"There is a plethora of different solutions in the context of big data. […] Unfortunately, not a single solution makes its users exempt from the need to deal with system programming. Different programs are needed depending on both characteristics of the data and characteristics of the architecture used for the analysis."* System programming, however, requires skills and expertise hardly available in organizations. Additionally, most technology providers introduced their own languages. The lack of a common language has not only hampered the emergence of querying and visualization tools in the past but also further increased the requirements imposed on data scientists.

'Establishing trust' was also considered challenging. In the context of big data, it happens easily that the staff in organizational departments cannot longer see the link between the processed data and the results of the analysis. The information systems researcher remarked, *"The staff in the organizational departments knows its field and data very well and gets irritated easily when results look strange and the analysis cannot be replicated."* Apart from the lack of transparency with respect to the analysis, doubts regarding the quality of the processed data undermine the establishment of trust. The databases researcher underlined the relevance of veracity for the establishment of trust, *"The origin of data is often questionable. Consequently, establishing trust is difficult. Inaccurate data implies inaccurate results."* It is not uncommon in the context of big data to analyze data collected

with other purposes in mind. It is only natural that data quality was not always a number one priority. The information systems researcher noted that the effort for data preparation is often too high.

'Assuring legal compliance' was considered another key challenge. One of the enterprise representatives gave an example from a research project, *"If data from two organizations is linked, which, in principle, can be very rewarding, data derivatives are generated. […] As there is no legislative framework in place defining to whom they belong, there is no alternative to deleting them."* In the context of big data, legal uncertainty is a topic not only with respect to data ownership but also with respect to data protection, insolvency and liability. National legislations often do not meet the needs of modern data processing. Apart from inadequacies in national legislations, difficulties in assuring legal compliance are often a consequence of transnational data processing, which is becoming increasingly important in practice but where ambiguity with respect to the applicability of possibly conflicting national legislations exists. The interviewee stressed with respect to data protection, *"The strong heterogeneity of data protection legislation in Europe is a key challenge. […] Adapting solutions to a large number of different legislations is hardly manageable. This represents a considerable locational disadvantage organizations in Europe face as opposed to ones in the US."*

B. Impact and Responses

The impact of the identified challenges, and thus their relevance, differs for organizations depending on their type, size and location. With respect to type, the interviewees differentiated between enterprises, bodies of the public administration and research institutions. Size was addressed in the interviews only with regard to enterprises, where a differentiation was made between small and medium-sized enterprises (SME), and large enterprises. Additionally, interviewees referred to highly specialized startups, which created their business model with an eye to big data. With respect to location, a differentiation was made between European and US organizations. For some of the challenges, the interviewees briefly sketched potential responses.

Identifying use cases was deemed particularly challenging for enterprises. As opposed to public administration bodies and research institutions, enterprises need to come up with compelling business cases. The information systems researcher stressed that it is essential for enterprises to show, on a reasonable efforts basis, that a use case is worthwhile. If this is not possible, priority is given to other projects. With respect to bodies of the public administration, the expert explained, *"They typically have large quantities of data and often are – keyword: open data – subject to a duty of disclosure or political pressure to make data available."* Not the calculation of a business case but the public interest is the decisive factor in this context. Research institutions interested in big data were deemed aware of the most promising use cases in their fields. Hiring competent experts and assuring legal compliance were considered particularly challenging for public administration bodies. One of the enterprise representatives explained that data scientists are typically paid very well. The interviewee feared that public administration bodies could be unable to pay competitive salaries. With respect to the assurance of legal compliance, one of the association representatives pointed out that, compared to enterprises and research institutions, public administration bodies are confronted with more comprehensive and complex legislative requirements, for instance, with respect to data protection. Hiring competent experts was deemed less relevant for research institutions. The reason given was that those, for which big data processing is relevant, most likely have the skills and expertise required. Furthermore, information literacy was deemed particularly prevalent in research institutions.

The interviewees considered setting up a technical architecture and hiring competent experts particularly challenging for SME with the exception of specialized startups. With respect to setting up a technical architecture, the experts pointed out that for SME there is usually no point in operating the compute clusters and storage systems required for the processing of big data. It was stated, however, that using shared infrastructures, as they are provided, for instance, within the scope of public or community cloud offers, could be an option to overcome the challenge. One of the interviewees representing an enterprise noted, *"SME normally cannot afford operating an own infrastructure. […] Based on cloud offers it is possible to flexibly use not only an infrastructure suitable for big data but also a fully-fledged big data architecture as a service."* According to the expert, cloud offers allow SME to experiment with the data they have access to. It was considered important, however, that potential users of cloud-based big data offers are aware of the inherent limitations and risks that come with using them. Just like public administration bodies, SME are also likely to be unable to pay the salaries data scientists expect. Additionally, hiring competent experts is particularly challenging for SME because data scientists often prefer to work for large enterprises. Promoting cooperation was deemed particularly challenging for large enterprises. The reason given was that in such enterprises it is more likely that use cases linking data from different organizational departments involve a large number of departments and people, which makes reaching consensus more difficult. One of the enterprise representatives suggested making a senior executive responsible for the promotion of cooperation among organizational departments.

For several reasons, setting up a technical architecture, carrying out advanced analytics and assuring legal compliance were considered more challenging for European organizations than for US organizations. Moreover, there are comparatively few European organizations on the market for big data technologies and services. The US were deemed approximately two years ahead in terms of big data by the interviewees. Most US organizations already have experience with big data. Also cloud offers are, according to the interviewees, more widely used in the US than in Europe. Consequently, setting up a technical architecture and carrying out advanced analytics is less challenging for them. To catch up, the interviewees suggested that Europe increases its efforts to create friendly conditions for entrepreneurship. Assuring legal compliance is more challenging in Europe than it is in

the US because the US legislation is more homogeneous. One of the enterprise representatives pointed out that in Europe more than two dozen different implementations of one solution are necessary to meet the national data protection legislations.

The interviewees did not mention any major differences in impact with respect to making solutions consumable and establishing trust. According to the interviewees, making big data solutions consumable implies bringing down the demands placed on the users. The databases researcher stressed accordingly, *"It is not an issue of training and education in the first place, it is a technical issue. […] To make the complex technologies consumable, we need a declarative language. We need an equivalent to SQL for big data."* Tasks formulated in such a language would then have to be compiled automatically in a way making optimum use of the given infrastructure. Apart from a declarative language, visualization tools were considered necessary to make the solutions usable by a larger group of staff members. The interviewees underlined that they consider user interfaces known from business intelligence (BI) solutions as useful starting points in this context. One of the interviewees representing an enterprise noted, *"There is a trend to reduce the demand placed on data scientists. It is increasingly tried to make solutions consumable by 'normal' business analysts, for instance, by 'harnessing' a BI user interface to a big data solution."* With respect to the establishment of trust, the relevance of both knowledge about big data and information literacy were emphasized.

IV. Discussion

In subsection A, the results from the series of interviews are collated into a coherent overall picture. Relevant findings from the quantitative studies, which served as input for the interviews, are provided in subsection B. Focused literature reviews allowed identifying the main directions in current research as well as promising avenues for future research. Both current directions and promising avenues are outlined in subsection C.

A. Overall Picture

When having a closer look at the challenges themselves, their impact and potential responses to them, it becomes obvious that they can well be categorized according to the information systems evaluation factors 'technology', 'organization' and 'people' [30, 31]. All but one of the challenges refer to one of the factors or any combination of the three. Fig. 1 illustrates the relations between the identified challenges and the three factors. It makes clear that the key challenges are not only technological in nature but that organizational and people-related matters as well as the legislative framework are also relevant.

The identification of use cases is an organizational matter, whereas carrying out advanced analytics is technological in nature. Hiring competent experts is a people-related matter. Setting up a technical architecture is both organizational and technological in nature as aspects of both areas have to be taken into account. An architecture has to meet the requirements resulting from an organization's specific big data use cases and respect technological feasibility. Promoting cooperation among organizational departments requires taking organizational and people-related matters into account. Making solutions consumable requires the consideration of both technological feasibility and individual user needs. Trust is at the center of the illustration as it is related to all three concepts. People have to trust in both technology and their colleagues. Building a culture of trust in the context of big data is an organizational matter. Assuring legal compliance is also related to all three concepts but, as opposed to establishing trust, it is, from an organizational perspective, much more outward oriented to the legislator and society in general.

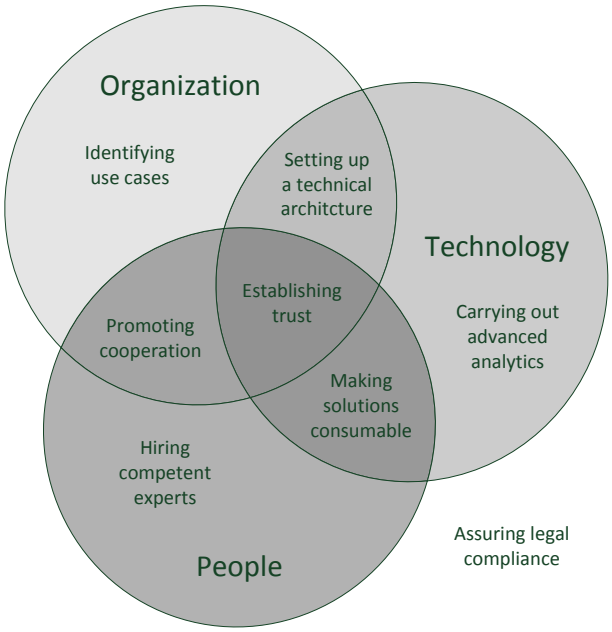

Fig. 1. Key challenges associated with big data

B. Quantitative Studies

With respect to some of the challenges, a closer look into the quantitative studies, which were also used for the preparation of the interviews and their analysis, brings to light additional insight.

The studies, for instance, underline not only the relevance of identifying use cases but also the importance of compelling business cases showing that big data use cases are worth investing [8, 7, 27, 29]. According to a TATA Consulting Services (TCS) study, enterprises find it difficult to understand where big data investments should be made, to determine what to do with the insight created from big data and to get the top management to approve investments in big data [8].

In addition to that, the studies reveal that determining which big data technologies to use is a difficult to obtain prerequisite for setting up a technical architecture [8, 27]. The TCS study indicates that big data technology is not yet considered mature and that it does not yet fulfill the hopes it

aroused. An Interxion study [28] stresses the role of legacy technology when building a technical big data architecture. Legacy technology does not only complicate setting up a technical architecture but also hampers carrying out advanced analytics. With respect to the implementation of big data analytics, a study conducted by The Data Warehousing Institute (TDWI) [27] found that scalability problems are among the top barriers. According to the TDWI study, current database software lacks in-database analytics, cannot process analytic queries fast enough, and cannot load data fast enough.

The promotion of cooperation between departments has not been addressed by many of the evaluated studies. The TCS study indicates, for instance, that getting business units to share information across organizational silos is critical to get business value from big data [8]. The evaluated quantitative studies make the breadth of problems related to hiring competent experts visible [7, 8, 27–29]. A Business Application Research Center (BARC) study [7], for instance, found inadequate technical and analytical know-how to have the highest degree of agreement among the study's participants. Reskilling the IT function to be able to use the new big data tools and getting the IT function to recognize that big data requires new technologies and skills were considered critical by the TSC study participants [8]. The study underlines that developing existing staff is an alternative to hiring new experts. Furthermore, the studies underline that putting the analysis of big data in a presentable form for making decisions is important [8, 7, 27, 28].

The participants of the TCS study [8] considered building high levels of trust between data scientists, who present insight on big data, and the functional managers, and getting functional managers to make decisions based on big data rather than on intuition critical. Both objectives are related to establishing trust in the context of big data. The former clearly refers to the trust between people in organizations, whereas the latter refers to trust in big data technologies and services. Concrete issues related to the assurance of legal compliance have hardly been touched on in quantitative studies. Data protection issues are, for instance, among the key challenges faced when processing big data according to the BARC study [7].

C. Research Directions

Academic literature on the identification of use cases for big data is rare. Bizer et al. [32], for instance, acknowledge that besides technological challenges, also making meaningful use of big data is a major challenge in the context of big data. There is plenty of literature, however, giving examples of how big data use cases could look like in enterprises [33], public administration bodies [34] or research institutions [35]. Segev et al. [36], for instance, explain how technology trends can be analyzed based on big data. Making the identification of use cases less challenging requires a combination of expert guidance, to compensate a potential lack of big data experience, and a systemic view of the organization, to ensure that the use cases create added value and encounter organizational demand. Descriptions of good practices and specific big data frameworks may thus be useful to facilitate the identification of big data use cases.

There is a considerable amount of academic literature on technical architectures for big data. Bakshi [37], for instance, reviews performance considerations and benchmarks. The article also addresses the role of solutions offered as public cloud services. Other pieces of related literature focus on specific solutions constituting big data architectures [38] or architectures tailored to specific fields of application [39]. To overcome the challenge inherent to setting up a technical architecture, it is probably most reasonable to take incremental steps and build on the existing capability to analyze data. Additionally, we consider the development and description of reference architectures a promising direction for future research.

There is plenty of literature on carrying out advanced analytics. The term 'advanced analytics', however, is interpreted very differently from one article to the other. Barton and Court [40], for instance, understand the term in its broadest sense and stress that advanced analytics is likely to become a decisive competitive asset in many industries. Articles focusing on carrying out specific clustering, classification or network analysis tasks on large amounts of heterogeneous data are comparatively rare. Cuzzocrea et al. [41], for instance, authored an article, which lists problems associated with both big data analytics in general and the processing of multidimensional data in particular. Overcoming the challenge inherent to carrying out advanced analytics requires substantial efforts by solution developers. The development of solutions natively supporting iterative algorithms on heterogeneous data is a key direction for future research.

The cooperation of departments and people in the context of big data has not received significant attention by the academic community in the past. McAfee and Brynjolfsson [6], for instance, noted in a general article on big data management that in the era of big data, organizations need to bring people who understand the problems together with the relevant data and people who have the skills and expertise required to effectively exploit the data. There is, however, plenty of literature on cooperation in organizations in general, which may be a promising starting point for future research focused on big data. Articles discuss, for instance, the role of social relations in general as foundation for cooperation [42] but also refer to the particular relevance of trust [43]. Making the promotion of cooperation between departments and people less challenging requires deep insight into aspects of organizational dynamics. As it is likely that the determinants of cooperation are largely consistent across organizations, we encourage related research. Approaches to overcome the challenge will benefit from progress in establishing trust.

With respect to hiring competent experts, the role of data scientists received a fair amount of attention. Davenport and Patil [44], for instance, provide a list of recommendations on how organizations can find experts meeting their specific needs. Others focus on specific demands on the experts and, for instance, highlight that data scientists have to feel comfortable speaking the language of business [6]. The academic community widely agrees that a large number of big data experts will be needed in the years to come [45]. The role of information literacy has been discussed mostly in the

context of education and librarianship [46, 47]. So far, the development of curricula for university programs and the role of organizational training programs focusing on big data have largely been neglected. Making the hiring of competent experts less challenging requires clarity with respect to the demands placed on them. Although it is up to the academic community to drive the development of suitable curricula, the wider availability of T-shaped professionals requires substantial commitment also by the organizations themselves. Overcoming the challenge will benefit from progress in making big data solutions consumable.

The 'consumability' of big data solutions by users has not yet received much attention in academic literature. Kandogan et al. [48] authored one of the few articles that discuss how the barriers for data-driven decision making can be lowered. They propose a self-service data intelligence system for business users, which does not only interpret user queries automatically but is also integrated with a data marketplace. According to Barton and Court [40], using a simple tool to deliver complex analytics has the potential to substantially reduce the need for new hires. Furthermore, it is agreed in the academic community that visualization tools increase in value in the context of big data [6, 49]. Overcoming the challenge inherent to making big data solutions consumable requires substantial efforts by solution developers. Users need to be exempt from the need to deal with system programming. Thus, the development of a declarative language, which is adopted across technology providers, is important. Industry associations may need to take over the lead in this regard.

Trust is essential because people can neither fully control the people in other departments or outside the organization who collect and share data, nor the technical architecture used for the analysis. Trust has not yet been made a topic in the context of big data often but insight from more general research on trust may be a promising starting point. So far, factors such as satisfaction with the existing system, task-technology fit and prior similar experiences were identified as determinants of technology trust [50]. To determine interpersonal trust, taking the complex web of existing and potential relationships into account is deemed important [51]. Making the establishment of trust less challenging requires insight into aspects of organizational dynamics. Investigating the specific determinants of trust in big data represents a worthwhile direction for future research.

Legal issues faced by organizations processing big data have received a fair amount of attention recently. Most of the attention has focused on data protection issues, though. Within the scope of the privacy literature, among others, the use of codes of conduct [52], access and transparency [53], and the right to be forgotten [54] were discussed. Others, such as Schadt [55] who writes about privacy in the life and biomedical sciences, focus on specific application areas of big data. None of them, however, tells organizations how to assure compliance with applicable data protection legislation. Wigan and Clarke [33] are among the few authors who go beyond data protection. They, for instance, refer to the notion of data ownership. Literature going into details with respect to ownership, insolvency or liability is still rare. Furthermore, legal uncertainty, which increases the difficulty of assuring legal compliance, has largely been neglected so far. Making the assurance of legal compliance less challenging requires a clear legislative framework, which takes the realities of modern data processing into account and is coordinated at an international level. Legal experts and policymakers are to take the lead in this regard.

V. Conclusion

The combination of a series of interviews with leading experts from enterprises, associations and research institutions, and focused literature reviews allowed identifying eight key challenges faced by organizations dealing with big data. The set of challenges includes identifying use cases, setting up a technical architecture, carrying out advanced analytics, promoting cooperation, hiring competent experts, making solutions consumable, establishing trust, and assuring legal compliance. The study shows that only some of the challenges are technological in nature; for others, organizational matters, or matters related to people and the legislative framework, respectively, are more relevant. For large enterprises, specialized startups, which created their business model with an eye to big data, and research institutions, it is typically easier to overcome the challenges than it is for other enterprises or bodies of the public administration. Although quantitative studies most widely underline the relevance of the challenges, they have not yet found their way on the agendas of the academic community to a considerable extent. Approaches to overcome the challenges are rare and promising directions for future research divers. However, as a thorough understanding of the challenges is a prerequisite for organizations to, first, make an informed decision for or against big data, and, second, if the decision is positive, overcome the challenges smoothly, research focusing on big data in general and related challenges in particular needs to be intensified. Artifacts that may be useful for organizations to overcome the addressed challenges include collections of good practices, frameworks specific to big data, technical reference architectures and a common declarative language to formulate tasks.

References

[1] S. Sicular, Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three 'V's. http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/ (accessed 2016/02/22).

[2] D. Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group, 2001.

[3] T. Lukoianova and V.L. Rubin, "Veracity roadmap: Is big data objective, truthful and credible?", Adv. Classif. Res. Online, vol. 24, no. 1, pp. 4–15, 2013.

[4] D. Beulke, Big Data Impacts Data Management: The 5 Vs of Big Data. http://davebeulke.com/big-data-impacts-data-management-the-five-vs-of-big-data/ (accessed 2016/02/22).

[5] S. Lohr, The Age of Big Data. http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html (accessed 2016/02/22).

[6] A. McAfee and E. Brynjolfsson, "Big data: The management revolution", Harvard Bus. Rev., vol. 90, no. 10, pp. 60–68, 2012.

[7] C. Bange, T. Grosser, and N. Janoschek, Big Data Survey Europe: Usage, Technology and Budgets in European Best-Practice Companies, Business Application Research Center (BARC), 2013.

[8] The Emerging Big Returns on Big Data: A TCS 2013 Global Trend Study, TATA Consulting Services (TCS), 2013.

[9] C. Lynch, "Big data: How do your data grow?", Nature, vol. 455, no. 7209, pp. 28–29, 2008.

[10] F. Frankel and R. Reid, "Big data: Distilling meaning from data", Nature, vol. 455, no. 7209, p. 30, 2008.

[11] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector", Commun. ACM, vol. 57, no. 3, pp. 78–85, 2014.

[12] E. Brynjolfsson and A. Saunders, Wired for Innovation: How Information Technology is Reshaping the Economy, Cambridge, MA: MIT Press, 2010.

[13] W.D. Eggers, R. Hamill, and A. Ali, "Data as the new currency: Government's role in facilitating the exchange", Deloitte Rev., no. 13, pp. 18–31, 2013.

[14] P. Hustinx, Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy, European Data Protection Supervisor, 2014.

[15] V. Mayer-Schönberger and K. Cukier, Big Data: A Revolution that will Transform How We Live, Work, and Think, Boston, MA: Houghton Mifflin Harcourt, 2013.

[16] E.W. Anderson, C. Fornell, and R.T. Rust, "Customer satisfaction, productivity, and profitability: Differences between goods and services", Market. Sci., vol. 16, no. 2, pp. 129–145, 1997.

[17] R. Oliva and R. Kallenberg, "Managing the transition from products to services", Int. J. Serv. Ind. Manag., vol. 14, no. 2, pp. 160–172, 2003.

[18] V. Gehlen, A. Finamore, M. Mellia, and M.M. Munafò, "Uncovering the big players of the Web", in Traffic Monitoring and Analysis, A. Pescapè, L. Salgarelli, and X. Dimitropoulos, Eds. Berlin: Springer, 2012.

[19] F. Schomm, F. Stahl, and G. Vossen, "Marketplaces for data", ACM SIGMOD Rec., vol. 42, no. 1, p. 15, 2013.

[20] A. Nadkarni and D. Vesset, Worldwide Big Data Technology and Services Forecast, 2015–2019, IDC, 2015.

[21] S. Kaisler, F. Armour, J.A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward", in Proceedings of the 46th Hawaii International Conference on System Sciences, Wailea, HI, 7-10 Jan. 2013.

[22] V. Marx, "Biology: The big challenges of big data", Nature, vol. 498, no. 7453, pp. 255–260, 2013.

[23] A. Jacobs, "The pathologies of big data", Commun. ACM, vol. 52, no. 8, pp. 36–44, 2009.

[24] D. Bollier, The Promise and Peril of Big Data, Washington D.C.: The Aspen Institute, 2010.

[25] A. Bogner and W. Menz, "The theory-generating expert interview: Epistemological interest, forms of knowledge, interaction", in Interviewing Experts, A. Bogner, B. Littig, and W. Menz, Eds. Basingstoke: Palgrave Macmillan, 2009.

[26] H.-F. Hsieh and S.E. Shannon, "Three approaches to qualitative content analysis", Qual. Health Res., vol. 15, no. 9, pp. 1277–1288, 2005.

[27] P. Russom, Big Data Analytics: TDWI Best Practices Report, The Data Warehousing Institute (TDWI) Research, 2011.

[28] Big Data - Beyond the Hype, Interxion, 2013.

[29] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, Analytics: The Real-World Use of Big Data: How Innovative Enterprises Extract Value from Uncertain Data, IBM, 2012.

[30] M.M. Yusof, A. Papazafeiropoulou, R.J. Paul, and L.K. Stergioulas, "Investigating evaluation frameworks for health information systems", Int. J. Med. Inform., vol. 77, no. 6, pp. 377–385, 2008.

[31] L.P. Willcocks, "Managing technology evaluation – techniques and processes", in Strategic Information Management: Challenges and Strategies in Managing Information Systems, R.D. Galliers and D.E. Leidner, Eds. New York, NY: Routledge, 2009.

[32] C. Bizer, P. Boncz, M.L. Brodie, and O. Erling, "The meaningful use of big data", ACM SIGMOD Rec., vol. 40, no. 4, pp. 56–60, 2012.

[33] M.R. Wigan and R. Clarke, "Big data's big unintended consequences", Computer, vol. 46, no. 6, pp. 46–53, 2013.

[34] J.C. Bertot and H. Choi, "Big data and e-government: Issues, policies, and recommendations", in Proceedings of the 14th Annual International Conference on Digital Government Research, Quebec City, QC, 17-20 Jun. 2013.

[35] C. Lynch, "Jim Gray's fourth paradigm and the construction of the scientific record", in The Fourth Paradigm: Data-Intensive Scientific Discovery, T. Hey, S. Tansley, K. Tolle, Hey, Anthony J. G, and K.M. Tolle, Eds. Redmond, WA: Microsoft Research, 2009.

[36] A. Segev, C. Jung, and S. Jung, "Analysis of technology trends based on big data", in Proceedings of the 2013 IEEE International Congress on Big Data, Santa Clara, CA, 6-9 Oct. 2013.

[37] K. Bakshi, "Considerations for big data: Architecture and approach", in Proceedings of the 2012 IEEE Aerospace Conference, Big Sky, MT, 3-10 Mar. 2012.

[38] P.C. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, New York, NY: McGraw-Hill, 2012.

[39] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, "Addressing big data issues in Scientific Data Infrastructure", in Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, San Diego, CA, 20-24 May 2013.

[40] D. Barton and D. Court, "Making advanced analytics work for you", Harvard Bus. Rev., vol. 90, no. 10, pp. 78–83, 2012.

[41] A. Cuzzocrea, I.-Y. Song, and K.C. Davis, "Analytics over large-scale multidimensional data", in Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP, Glasgow, 28 Oct. 2011.

[42] G. Spagnolo, "Social relations and cooperation in organizations", J. Econ. Behav. Organ., vol. 38, no. 1, pp. 1–25, 1999.

[43] D.J. McAllister, "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations", Acad. Manag. J., vol. 38, no. 1, pp. 24–59, 1995.

[44] T.H. Davenport and D.J. Patil, "Data scientist: The sexiest job of the 21st century", Harvard Bus. Rev., vol. 90, no. 10, pp. 70–76, 2012.

[45] J. Manyika, M. Chui, B. Brown, J. Bughin, and R. Dobbs, et al., Big data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey & Company, 2011.

[46] J. Carlson, M. Fosmire, C.C. Miller, and M.S. Nelson, "Determining data information literacy needs: A study of students and research faculty", Portal Libr. Acad., vol. 11, no. 2, pp. 629–657, 2011.

[47] N.M. Weber, C.L. Palmer, and T.C. Chao, "Current trends and future directions in data curation research and education", J Web Libr., vol. 6, no. 4, pp. 305–320, 2012.

[48] E. Kandogan, M. Roth, C. Kieliszewski, F. Ozcan, and B. Schloss, et al., "Data for all: A systems approach to accelerate the path from data to insight", in Proceedings of the 2013 IEEE International Congress on Big Data, Santa Clara, CA, 6-9 Oct. 2013.

[49] S. LaValle, E. Lesser, R. Shockley, M.S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value", Sloan Manag. Rev., vol. 52, no. 2, pp. 21–31, 2011.

[50] S.K. Lippert and H. Forman, "A supply chain study of technology trust and antecedents to technology internalization consequences", Int. J. Phys. Distrib. Logist. Manag., vol. 36, no. 4, pp. 271–288, 2006.

[51] D.L. Ferrin, K.T. Dirks, and P.P. Shah, "Direct and indirect effects of third-party relationships on interpersonal trust", J. Appl. Psychol., vol. 91, no. 4, pp. 870–883, 2006.

[52] I.S. Rubinstein, "Big data: The end of privacy or a new beginning?", Int. Data Priv. Law., vol. 3, no. 2, pp. 74–87, 2013.

[53] O. Tene and J. Polonetsky, "Big data for all: Privacy and user control in the age of analytics", Nw. J. Tech. Intell. Prop., vol. 11, no. 5, 2013.

[54] B.-J. Koops, "Forgetting footprints, shunning shadows: A critical analysis of the 'right to be forgotten' in big data practice", SCRIPTed, vol. 8, no. 3, pp. 229–256, 2011.

[55] E.E. Schadt, "The changing privacy landscape in the era of big data", Mol. Syst. Biol., vol. 8, 2012.