The Macrotheme Review

A multidisciplinary journal of global macro trends

BIG DATA CHARACTERISTICS

Peter Géczy
National Institute of Advanced Industrial Science and Technology (AIST), Japan

Abstract

Explosion of digital data and its diversity over the past few decades has attracted significant attention. Complex systems and devices are capable of generating vast amounts of serviceable data. The data reflects various operational aspects such as functionality of systems, interactivity of humans with digital devices and environments, and responsivity to internal and external stimuli from sensors. Big, diverse and rapidly produced data presents various novel challenges, but also opportunities. Big data allows us to tackle longstanding complex problems. It also provides opportunities to explore new scientific domains that have only recently emerged—thanks to availability of data. However, big data also highlights novel challenges ranging from acquisition and technological issues, throughout processing and maintenance, to business and social consequences. Big data trend, despite strongly emerging in several domains, has been lacking appropriate elucidation. We attempt to explore pertinent interdisciplinary characteristics of big data at the intersections of its technological and operational enablers.

Keywords: big data characteristics, data aspects, processing capabilities, analytics, actionable knowledge, information technology management, management of information systems

1. INTRODUCTION

Contemporary organizations significantly rely on a broad range of information technologies. Many organizations would be unable to function without the operational support provided by information technologies (Turban and Volonino, 2011). Knowledge-intensive organizations designate information technologies among their core assets (Alvesson, 2004). Knowledge workers increasingly depend on information systems and services (Davenport, 2005). They often incorporate the essential business processes. Majority of former pen-and-paper business processes have been transferred into digital forms. Such transformation facilitates improved working efficiency, productivity, task automation, as well as accessibility of information, documents and resources (Wikoff, 2008).

Information technology adoption pathways have been varying—depending on organization. However, there is a noticeable pattern. In early days, organizations had their own strategies for building information technology capabilities, resources and infrastructures. In this period, dedicated information technology departments have been relatively absent. Lack of experience with information technologies has been reflected in absence of best practices. There was also lack of coordinated long-term strategy and planning (Butler and Murphy, 2007). Departments within organizations have been implementing their own information infrastructures and systems (Papastathopoulou et al., 2007).

A need for coordinating strategy, planning and deployment of information technologies within organizations has emerged (Georgantzas and Katsamakas, 2010; Boar, 2000). However, radical reengineering of deployed systems would be costly and could hinder operations in organizations. Solutions that could employ existing technologies have been favored. One viable solution has been a deployment of organizational portals (Collins, 2000). The portals have provided single-point access to systems and services distributed over various departments (Oertel et al., 2010; Sullivan, 2004). Enabling technologies have been standardized network communication protocols, web, and service-oriented architecture and design (Rosen et al., 2008). Front-ends of portals have been web-based systems, while back-ends have been database and legacy technologies. Service-oriented approaches provided interoperability of front-ends and back-ends.

Internet and web technologies allowed global interconnectivity and access of resources (Knight, 1998). Many resources have been digitalized and made available. Digitalization of existing analog resources and media has been one of the first waves of digital data expansion. Another wave has been a vast amount of content produced by organizations and individuals (Krumm et al., 2008). Broad spectrum of content has chiefly contributed to diversity of digital data. Expanding content and services available on the web have attracted large numbers of users. Businesses have started tracking users and analyzing data about their interactions (Lackner et al., 2010).

Web servers feature data logging capabilities that paved the way for analyzing users' web interactions (Geczy et al., 2007 and 2008). Log data analysis provides a reasonable level of detail on system functions and users' interactions with portals and services (Kaushik, 2009). However, log data grows rapidly and requires substantial processing. Hence, other data collection technologies have been developed. Data collection technologies expanded to specialized hardware tools allowing deep inspection of communication packets and system functions. While advanced data acquisition technologies are valuable, they also raise concerns about security and privacy (Anthes, 2010; Lanois, 2010).

Data has become a currency of information economy (St. Amant and Ulijn, 2009). Value of data has been rising and organizations have been realizing it (Lievesley et al., 1993). They have started collecting their own data and exploring various monetization opportunities. Sale of collected data has become a viable revenue stream. However, organizations have been realizing prospects not only in selling data, but also in analyzing it with the aim of improving their own operations. Hence, organizations have kept on expanding their data collecting and processing methods (Davenport et al., 2007 and 2010). Expansion of data acquisition operations has led to rapid growth of data and processing demands (Frischbier and Petrov, 2010). Consequently, novel challenges have emerged—big data problems (Buhl et al, 2013; Hunter, 2013; Klein et al., 2013; Walsh et al., 2012).

Big data problems have arisen from disproportionate growths between collected data and capabilities of organizations to process it. Data has been growing considerably faster than advances in processing technologies. Big data issues have become significant. They require interdisciplinary elucidation in addition to advances in related individual disciplines. Despite the importance of the problem, there is an absence of interdisciplinary studies. This work attempts to fill the gap by exploring characteristics of big data and providing an encompassing perspective.

2. BIG DATA PROBLEM

Organizations have been routinely dealing with various data in their operations. Although they have been processing data for various purposes, the problem of big data has emerged only relatively recently (Buhl et al, 2013; Hunter, 2013; Klein et al., 2013; Walsh et al., 2012). There are three main contributing factors leading to emergence of big data problems in organizations: digitalization of business processes and associated data, accumulation of additional data, and aspirations to extract actionable knowledge from data and/or monetize data.

Digitization of business processes has enabled organizations to transfer formerly paper-based processes into digital business processes, environments and platforms. Business process flows have been reengineered and implemented into organizational internal digital environments. Together with business processes, associated data has also been digitized (Oertel et al., 2010). Digitization of business processes and data have allowed organizations to advance in automation and innovation of business processes (Beydoun, 2013). Digitization has improved operating efficiency of organizations, but also increased their reliance on information technologies.

Acquisition of additional digital data related to organizations' operations have been progressing alongside with business process digitization. Organizations have been realizing that acquired data has beneficial uses (Geczy at al., 2011). For instance, business processes on organizational internal portals have implemented technologies for collecting data about users' interactions with the digital environments. The data has enabled monitoring of usability of deployed processes and services and targeting of improvements and innovations of services at organizational portals. Hence, organizations have been implementing technologies enabling them to acquire expanding volumes of data.

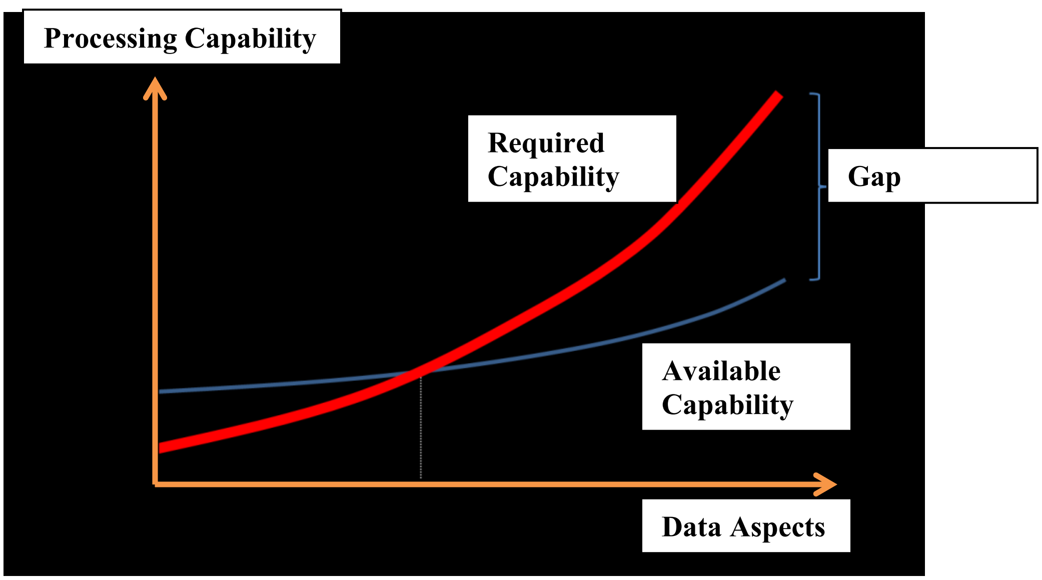

Figure 1. *Illustration of the big data problem definition in organizations. The problem of big data is a reflection of discrepancy between the available capabilities to process acquired data and the required capabilities given the contemporary state of processing methods.*

Having extensive amounts of data, organizations have been exploring opportunities to utilize the accumulated data—both internally and externally (Davenport et al., 2007 and 2010). Internal data utilization has targeted improving core competencies, increasing operational efficiency, and innovating business processes. External data utilization has aimed at monetizing the data by selling it to third parties; hence, establishing additional revenue streams for organizations. Some data can be utilized directly without extra processing; however, majority of the data requires additional processing. Data processing has targeted two main goals: preprocessing and formatting of data for sale to external purchasers, and analyzing data and extracting actionable knowledge (Laursen and Thorlund, 2010). Both targets necessitate addressing various data related issues.

Accumulation and processing of large amounts of data by organizations have led to numerous challenges (Wigan and Clarke, 2013; Buhl et al, 2013; Klein et al., 2013; Hunter, 2013; Walsh et al., 2012). These challenges are commonly denoted as big data problems. It is important to note that big data problems do not relate only to the issues arising from the data size. Information technology providers doing business in 'big data' often propagate such meme. Understandably, such meme serves their marketing and business purposes. However, this is a rather narrow view.

We approach the big data problem from the holistic relational perspective—as a relation between various aspects of data and its processing. In such approach, the size of data is just one of the aspects of data. The significant advantage of this relational approach is that it enables encompassing and clarifying interrelationships between the data and its processing. The problem of big data, in the presented relational approach, is illustrated in Figure 1. Horizontal axis represents various aspects of data, such as volume, diversity, rate at which data is acquired, etc. Vertical axis represents processing capability—including processing power organizations have at their disposal. The big data problem is defined as a gap between required capability to process data and organization's available capability.

For instance, assume that organization needs to process 1GB of data every second. If processing of 1KB of data requires ten-thousand processing instructions (given the available and implemented algorithms), to process 1GB of data per second requires capability to process ten-billion instructions per second. Hence, to achieve the desired outcome, organization must have available computing power of at least ten-billion instructions per second. Having any less computing power would render the organization incapable to meet its data processing requirements. In other words, the organization would be facing the big data problem. To solve the problem, there are several options. One option would be to increase its computing power, e.g. by purchasing new servers, or upgrading the existing ones. Another option would be to reduce the amount of data, e.g. by deploying data filtering methods.

Expressing the big data problem as a relationship between the aspects of data and its related processing capabilities has various benefits. It permits clear distinctions between processing aspects and data aspects. Furthermore, it permits observations of their interrelationships. Clear distinctions allow in-depth elucidation of both data aspects and processing aspects. Organizations can explore the specific issues they are facing with their data and its processing. They can accurately evaluate elements in their respective domains. Interrelationships highlight the links between data aspects and processing aspects. This allows observing various influences between interconnected aspects. In other words, how changes in certain aspects of data influence changes in certain aspects of processing–and vice versa. Having such encompassing perspective enables organizations to accurately model and optimize deployment and innovation of existing resources as well as plan acquisitions of future resources.

Following the trend of collecting, processing and monetizing data exposes organizations to numerous issues encompassed in the above mentioned big data problem. There are three main categories of issues that organizations face. These issues are related to data, its processing and management. Data related issues pertain to aspects of data and its acquisition methods. Processing related issues reflect feasibility challenges of analytic targets due to computational demands, algorithms and methods. Management related issues are concerned with managing infrastructures, computational architectures and innovation, in order to meet the demands of organizations. The following sections address pertinent aspects related to data, processing and management.

3. DATA RELATED ASPECTS

Organizations collect data from both internal and external sources. Internal sources provide data related to their internal operations and business processes. External sources supply data from their externally oriented business operations with business partners, suppliers, customers, etc. Data originating from both sources has various aspects that play significant roles in their acquisition, processing and management. Among the most important characteristics of data is its sensitivity that determines its further handling. Several other data characteristics are also important, such as volume, diversity, and quality. The following paragraphs elaborate on these aspects.

Sensitivity. Data sensitivity is the crucial aspect for majority of organizations. It expresses whether data contains sensitive information such as personally identifiable information, confidential internal information, nondisclosure information, etc. Sensitivity determines data handling requirements. For instance, many countries have legislative frameworks that regulate collection, handling, and management of such data. If data contains personally identifiable information, organizations must comply with the regulatory requirements. Analogously, if data contains confidential internal information (e.g. internal know-how or financial data), it is in organization's best interest to properly protected it against external exposure and compromisation. Such data must be stored in-house and not on external cloud providers. Furthermore, its internal exposure must be protected by measures such as encryption and its access must be restricted only to relevant members of the organization.

Diversity. Diversity of data refers to the spectrum of different element types within the data. Data can be significantly diverse. For instance, mobile devices such as smart phones are able to produce various types of data; e.g. audio, video or multimedia data, text data, location data, temporal data, gyroscopic data. Different data types generally require different handling and impose diverse requirements on allocation capacity, speed of processing and other issues. While data diversity may be beneficial in some cases, it may be detrimental in others. Data diversity adds richness, but it also adds complexities to processing and maintenance. Organizations must account for various conditions and requirements related to data diversity.

Quality. Qualitative aspects of data are notably important. Various elements of data may have different types and quality characteristics. There are various features affecting data quality; such as completeness, accuracy, timelines, etc. Completeness of data indicates whether there are missing values in data. Data without missing values is complete. While completeness is easy to achieve for some data, it may be difficult for others. Missing values in data may be due various causes; such as unavailability or undetectability. Accuracy of data underlines the precision with which the data is expressed. For instance, a timestamp on a transaction may be accurate up to a second, while date of birth is accurate up to a day. Organizations should establish their own qualitative metrics for data, or adopt commonly accepted metrics.

Volume. Data volume generally refers to its size in terms of standard information metrics (bits and bytes). The meme *big data* is commonly associated with this aspect of data. The volume of data used to play important role in storage and processing. However, the storage capacity issue has become less pressing due to rapidly progressing storage technologies and decreasing prices per gigabyte. Contemporary storage technologies permit implementations of efficient cost-performance storage solutions for internal deployments. Organizations can also outsource their storage needs for non-sensitive data to external cloud-based providers. Data volume, however, plays major role in processing considerations. Although the computational power of processing systems has been steadily increasing, the data volumes have been rising faster.

Speed. Data speed has two components: inflow and outflow speeds. If both inflow and outflow speeds are equal, then single data speed refers to both of them. Inflow speed of data denotes either how fast is data acquired, or the maximum speed of the input channel. Analogously, outflow speed of data denotes either how fast data flows out of a system, or the maximum speed of the output channel. Data speed is usually measured in bits or bytes per second. Different data types may require different data speeds. For instance, while collection of audio data may require speeds in the range of hundreds of kilobits per second, video data usually require megabits per second.

Structure. Data structure underlines the degree of its organization. Structured data has a high degree of organization, whereas unstructured data lacks sufficient degree of organization. Unstructured data is more suitable for human processing—for instance, text document or e-mail. Structured data is suitable for machine processing. If all data were properly structured, it would provide significant advantages for computer processing. Structured data can be seamlessly included into conventional databases and managed with database tools. Unstructured data is generally more difficult to process since it usually requires pre-processing before conventional tools can process it.

4. PROCESSING RELATED ASPECTS

Data processing is among the core concerns of organizations collecting large volumes of diverse data. Unprocessed data has relatively limited value. Since it is important for data holders to increase and derive additional values from accumulated data, the data processing plays the key role. Two main categories of processing needs are commonly distinguished: data management processing and analytics. Data management processing aims at preparing data for its efficient management and future processing. Analytics aim at analyzing data in order to extract insights and actionable knowledge. Extracted actionable knowledge has beneficial uses for organizations. Although targets of data management processing and analytics largely vary, there are several common challenges associated with big data processing. Selected pertinent aspects are addressed in the following paragraphs.

Available Algorithms. At the heart of every computerized processing are algorithms. Despite the significant progress in computer science, data mining, analysis and other disciplines over the past decades, algorithmic processing is suitable for relatively limited set of tasks. Computerized processing is generally well suited for tasks requiring repetitive mathematical calculations and logical operations. However, the tasks similar to higher-order human cognitive processing are still largely out of reach for machines. Therefore, it is imperative to have realistic expectations about what can be achieved with the implemented algorithms and collected data.

Scalability. Scalable processing systems are able to flexibly deal with expanding data. That is, either they can meet the processing demands resulting from growing data, or they can be suitably expanded to meet the demands. For majority of organizations a suitable expansion is linear—where resource demands are linearly proportional to data growth. Linear scalability should be reflected in both computing power and financial resources. For instance, if linearly increasing computing power results in exponential cost growth, then eventually processing systems become too costly—organization may not be able to afford processing of big data.

Timelines. Timelines encompasses temporal requirements for providing processing results. Two major processing modes are ordinarily used for providing timely results: online and offline processing. Online mode refers to immediate processing of data. Offline mode denotes data processing after some delay. Online processing is used in real time systems—where data must be processed as soon as it becomes available. For example, data generated by avionic sensors on a plane during a flight requires immediate processing. Online and offline processing modes impose different requirements on computing power. Big data needing immediate processing requires considerably greater computing power than data requiring processing during three days.

Analytic Targets. Targeting of suitable analytics is an important aspect influencing data acquisition, processing and management (Bernhardt, 2004). While academic and research organizations may be serendipitous, business organizations usually demand rigorous planning and optimization of resources.

Proper planning of analytic targets enables determining suitable data collection, processing and management methods. For instance, if web-based shop wants to measure conversion rates, i.e. how many web site visitors buy something, it needs to collect only data on unique visitors and buyers. It does not need to collect data about what each visitor clicked on the web site. Such data would only increase data storage and processing demands.

Actionable Knowledge. Deriving actionable knowledge from collected and analyzed data is the primary objective of many organizations. However, actionable knowledge is subjective. Managers in one organization may consider outcome of some analytics to be actionable knowledge, however, managers in another organization may consider it insufficient for making decisions. Subjectivity of actionable knowledge demands tailored approach. Processing and analytic targets should be sufficiently tailored to meet the required demands for actionable knowledge by the actual users in organizations. One-fit-all solutions should be avoided.

4. MANAGEMENT RELATED ASPECTS

Management aspects arising from the big data problem encompass both data collection and processing domains. Acquisition and processing of large data volumes present novel managerial challenges for organizations (Tallon, 2013). Organizations need to adopt new management approaches that may significantly depart from the methods they have been using (Malik, 2013). Adoption of new management methods goes alongside the deployment of novel technologies and/or reengineering of the existing information technology infrastructures. Such changes may require significant investments. While bigger organizations can afford larger investments and faster adoption, smaller organizations may decide to transitions more slowly. Essential managerial aspects are described in the following paragraphs.

Infrastructure Reengineering. Big data problems affect information technology infrastructures. Capability to collect and process big data necessitates sufficient transmission and storage capacities, as well as computing power. Large internal data flows demand fast intranet connectivity, and external data flows fast internet connectivity. External internet service providers usually provide internet connectivity. Choice of internet service providers does not generally affect communication infrastructures within organizations. However, intranet connectivity is under control of organizations and may require significant changes.

Resource Optimization. Challenges arising from managing big data require proper optimization of resources. Big data affect human, financial and physical resources. Adoption of novel technologies may require personnel changes and retraining. This takes time and investments. Deployment of physical resources such as storage, computing and communication equipment also requires time and investments. To avoid high initial costs and adoption issues, it is advisable to balance the need for new resources with the utilization of the existing ones. Cloud computing architectures allow such balancing (Geczy et al., 2012). Existing organizational resources should be optimized in order to manage transition at desirable pace and affordable cost.

Storage Management. Big data volumes impose additional demands on storage resources. Rapid progress in storage technologies is pushing per-megabyte costs continually down and storage density up. Hence, the hardware issues are gradually becoming less concerning. On the other hand, architectural issues are becoming dominant. Properly implemented storage architecture requires at least three stages. The primary storage (1st stage) focuses on fast access. The fastest storage units are internal memories, but they have relatively small capacities. The secondary storage (2nd stage) balances the speed and capacity demands―achievable by solid state drives. The tertiary storage (3rd stage) focuses on large capacities and backups―utilizing conventional hard drives and Blu-ray discs.

Databases. Database technologies are used for efficient management and processing of big data. Traditionally, large databases have been software systems running on specialized single-rack high-performance hardware. These approaches have been unable to keep up with rapid growth of data and processing demands. Specialized hardware systems have become exceedingly expensive. This has given rise to technologies employing large quantities of commodity computing hardware, such as personal computers. Big data volumes and computing power can be distributed over large numbers of connected computers. Large numbers of connected machines–forming clusters–are able to deal efficiently with big data storage and processing. Diversity of big data requires deployment of both SQL and NoSQL database systems. SQL systems are more suitable for structured data whereas NoSQL databases are more suitable for unstructured data.

Legal Aspects. Organizations are required to comply with legal requirements that regulate some data. Handling of sensitive data such as personal information is strictly regulated in many countries. Legislations regulate various aspects of data collection, processing, retention and exposure (Joseph and Johnson, 2013). Organizations collecting, storing or processing such data must meet the legal requirements in countries where they operate. That is, they must implement systems, measures and management practices that are in accordance with effective legislations. Managing compliance with legal requirements is important and necessitates knowledge of legal frameworks.

5. CONCLUSIONS

Exponential expansion of generated digital data is disproportional to the progress in data processing and storage technologies. This leads to a situation when available data is excessively big, and organizations are unable to adequately process and manage it. Inability to cope with the aspects of growing data has given rise to newly emerging challenges—big data problems. The presented approach to the big data problem underscores the relationship between the aspects of data and processing capabilities. Data volume is just one of the aspects; however, several other data aspects play significant roles. The big data problem is delineated as a gap between required and available capabilities of organizations to satisfactorily deal with the aspects of collected data in its processing and management. This relational formulation of the big data problem highlights interconnectedness of three main sets of aspects; that is, aspects of data, processing and management.

Data aspects encompass both qualitative and quantitative characteristics of data and its acquisition. Notable qualitative aspects of data include sensitivity, quality and structure. Sensitivity of data relates to impacts of potential data exposure, damage, of various forms of compromisation. Sensitive data requires proper handling, security and compliance with established legal frameworks where organizations operate. Quality of data denotes aspects such as missing values and accuracy. These aspects have also impact on qualitative features of processing. Pertinent quantitative aspects of data are size and diversity. Voluminous and diverse data pose processing and storage challenges. Processing of large volumes of diverse data requires significant computing power. Rising demands on computing power are not the only challenges associated with big data processing. Suitable processing of data depends on viability of analytic targets, availability of efficient algorithms and their scalability. It is unreasonable to aim at analytic targets for which there are nonexistent or unscalable algorithmic methods. Contemporary repertoire of efficient and scalable algorithms and tools addresses relatively unsophisticated tasks. Organizations must carefully weigh whether collection and processing of big and diverse data would lead to justifiable benefits in a long term. They must manage and optimize their resources accordingly. Management aspects related to big data often include infrastructure reengineering, storage management and legal issues. Information technology infrastructures in many organizations must undergo significant changes, to be able to store and process collected data. They must also meet the associated legal requirements. Properly managing transition to utilizing big data and analytics is crucial.

REFERENCES

Alvesson, M. (2004). *Knowledge Work and Knowledge-Intensive Firms.* Oxford University Press, Oxford.

Anthes, G. (2010). Security in the Cloud: Cloud Computing Offers Many Advantages, but Also Involves Security Risks. *Communications of ACM*, **53**(11), 16-18.

Bernhardt, V. (2004). *Data Analysis (2nd Ed.)*. Eye on Education, Larchmont.

Beydoun, G. (2013). Dynamic Evaluation of the Development Process of Knowledge-based Information Systems, *Knowledge and Information Systems*, **35**(1), 233-247.

Boar, B. H. (2000). *The Art of Strategic Planning for Information Technology (2nd Ed.)*. Wiley, Indianapolis.

Buhl, H. U. Raglinger, M. Moser, F. Heidemann, J. (2013). Big Data, *Business & Information Systems Engineering*, **5**(2), 65-69.

Butler, T., Murphy, C. (2007). Understanding the Design of Information Technologies for Knowledge Management in Organizations: A Pragmatic Perspective. *Information Systems Journal*, **17**(2), 143-163.

Collins, H. (2000). *Corporate Portals: Revolutionizing Information Access to Increase Productivity and Drive the Bottom Line*. Amacom, New York.

Davenport, T. H. (2005). *Thinking for a Living - How to Get Better Performance and Results from Knowledge Workers.* Harvard Business School Press, Boston.

Davenport, T. H., Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business School Press, Boston.

Davenport, T. H., Harris, J. G., Morison, R. (2010). *Analytics at Work: Smarter Decisions, Better Results*. Harvard Business School Press, Boston.

Frischbier, S., Petrov, I. (2010). Aspects of Data-Intensive Cloud Computing. *LNCS 6462*, Springer-Verlag, 57-77.

Géczy, P., Izumi, N., Akaho, S., Hasida, K. (2007), Knowledge Worker Intranet Behaviour and Usability. *International Journal of Business Intelligence and Data Mining*, 2(4), 447-470.

Géczy, P., Izumi, N., Akaho, S., Hasida, K. (2008), Enterprise Web Services and Elements of Human Interactions. *Business Information Systems* (W. Abramowicz and D. Fensel, Eds.), Springer-Verlag, 263-272.
Géczy, P., Izumi, N., Hasida, K. (2012). Cloudsourcing: Managing Cloud Adoption. *Global Journal of Business Research*, **6**(2), 57-70.

Géczy, P., Izumi, N., Hasida, K. (2011). Foundations for Effective Portal Service Management. *Global Journal of Business Research*, 5(2), 131-141.

Georgantzas, N. C., Katsamakas, E. G. (2010). Performance Effects of Information Systems Integration: A System Dynamics Study in a Media Firm. *Business Process Management Journal*, **16**(5), 822-846.

Hunter, M. G., (2007). *Contemporary Chief Information Officers: Management Experiences*. IGI Publishing, Hershey.

Hunter, P. (2013). Journey to the Centre of Big Bata. *Engineering & Technology*, **8**(3), 56-59.

Joseph, R.C., Johnson, N.A. (2013). Big Data and Transformational Government. *IT Professional*, 15(6), 43-48.

Kaushik, A. (2009). *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*. Sybex, Indianapolis.

Klein, D. Tran-Gia, P. Hartmann, M. (2013). Big Data. *Informatik-Spektrum*, **36**(3), 319-323.

Knight, F. S. (1998). The Internet and Web Go Mainstream. *Business Communications Review*, 28(10), 42-46.

Krumm, J. Davies, N. Narayanaswami, C. (2008). User-Generated Content. *IEEE Pervasive Computing*, **7**(4), 10-11.

Lackner, G. Teufl, P. Weinberger, R. Heng, S.-H. Wright, R.N. Goi, B.-M. (2010). User Tracking Based on Behavioral Fingerprints. Journal on Data Semantics, **6474**, 76-95.

Lanois, P. (2010). Caught in the Clouds: The Web 2.0, Cloud Computing, and Privacy? *Northwestern Journal of Technology and Intellectual Property*, **9**(2), 29-49.

Laursen, G. H. N., Thorlund, J. (2010). *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*. Wiley, Indianapolis.

Lievesley, D. Ross, S. Higgs, E. (1993). Increasing the Value of Data. *BLRD Reports*, **6112**, 205-218.

Malik, P. (2013). Governing Big Data: Principles and Practices. *IBM Journal of Research and Development*, 57(3/4), 1-13.

Oertel, N., Dibbern, J., Nochta, Z. (2010). Assessing the Potential of Ubiquitous Computing for Improving Business Process Performance. *Information Systems and e-Business Management*, **8**(4), 415-438.

Papastathopoulou, P., Avlonitis, G. J., Panagopoulos, N. G. (2007). Intraorganizational Information and Communication Technology Diffusion: Implications for Industrial Sellers and Buyers, *Industrial Marketing Management*, **36**(3), 322-336.

Rosen, M., Lublinsky, B., Smith, K. T., Balcer, M. J. (2008). *Applied SOA: Service-Oriented Architecture and Design Strategies*. Wiley, Indianapolis.

Sullivan, D. (2004). *Proven Portals: Best Practices for Planning, Designing, and Developing Enterprise Portal*. Addison-Wesley, Boston.

St. Amant, K. Ulijn, J.M. (2009). Examining the Information Economy: Exploring the Overlap between Professional Communication Activities and Information-Management Practices. *IEEE Transactions on Professional Communication*, **52**(3), 225-228.

Tallon, P.P. (2013). Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. *Computer*, **46**(6), 32-38.

Turban, E., Volonino, L. (2011). *Information Technology for Management: Improving Strategic and Operational Performance (8th Ed.)*. Wiley, Indianapolis.

Walsh, R. O Callaghan, R. Yoffou, S. Hughes, R. (2012). Big Data Is a Solution - So Where's the Problem?, Cutter IT Journal, **25**(10), 6-12.

Wigan, M.R., Clarke, R. (2013). Big Data's Big Unintended Consequences. *Computer*, **46**(6), 46-53.

Wikoff, L. (2008). The Importance of Process in IT Organizations. *Journal of Computing in Civil Engineering*, 22(2), 71-73.