Journal of Business Research 128 (2021) 381–390

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Big data analytics for sustainable cities: An information triangulation study of hazardous materials transportation

Lisha Ye[a], Shan L Pan[b], Jingyuan Wang[c,f,*], Junjie Wu[d], Xiaoying Dong[e]

[a] Department of Strategy and Industry, China Mobile Research Institute, China
[b] School of Information Systems, Technology and Management, UNSW Business School | UNSW, Australia
[c] School of Computer Science and Engineering, Beihang University, China
[d] School of Economics and Management, Beihang University, China
[e] Guanghua School of Management, Peking University, China
[f] State Key Laboratory of Software Development Environment, Beihang University, Beijing, China

ARTICLE INFO

Keywords:
Big data analytics
Grand challenges
Information triangulation
Case study

ABSTRACT

Big data analytics (BDA) is regarded as an advanced tool for achieving sustainable development as part of the grand challenges (GCs). However, it is not clear how BDA can be used by data scientists to solve the GCs with multisource data in a cross-disciplinary approach. Based on a case study of city-based dangerous goods transportation (DGT), this paper explores how data scientists use BDA to triangulate data, methods, knowledge and solutions for solving GCs. The contribution of this study is threefold: (1) it contributes to research on GCs and discusses how BDA can be used in problem solving for multidomain GCs from a management perspective; (2) it enriches the theory of information triangulation and proposes several steps for information triangulation in BDA to solve GCs; and (3) it contributes some practical implications for the management of organizations when solving social problems and pursuing sustainable development.

1. Introduction

Grand challenges (GCs) are highly significant societal problems that can be plausibly addressed through coordinated and sustained effort (Eisenhardt et al., 2016; George et al., 2016), such as climate change, water scarcity, healthcare provision, and poverty alleviation (Ferraro et al., 2015). Solutions to GCs can generate an enormous global impact on natural ecology and human activities. Traditionally, GCs belong to a specific scientific field and can be solved by domain-specific knowledge, which can generate breakthroughs in zoology, meteorology, physics or other disciplines. However, in this era, there are also many GCs involving diverse domains that require multidata sources, interdisciplinary knowledge and coordinated efforts.

With the development of science technologies, digital sustainability is gradually becoming a goal for organizations aiming to solve multidomain GCs and achieve sustainable development by leveraging technologies (George et al., 2019; Tim et al., 2018). Big data analytics (BDA) is an advanced analysis and processing tool for realizing digital sustainability (Chen et al., 2012). Big data (BD) refers to large and varied data from multiple sources, and BDA, the application of statistical, processing and analytics techniques to BD, is used to mine for complex hidden information and knowledge (George et al., 2016; Grover et al., 2018). In the context of GCs, BD can be collected from various channels, such as sensors, devices, and people, and thereby present a comprehensive reflection of a specific social problem (Grover et al., 2018). BDA is a powerful data processing tool, and users can benefit from the ability to derive insights and make decisions based on multiple sources of BD to address GCs (Dremel et al., 2018).

Existing studies on BDA focus more on the technology level, including analysis tools, processing, challenges and potential applications (Agarwal & Dhar, 2014; Chen et al., 2012; Günther et al., 2017). However, tackling GCs is not only a technical problem but also a managerial problem (George et al., 2016). For example, solving GCs may involve determining how to identify and seek problem-related data, how to compare different analysis methods and choose the best, and how to work across multiple disciplines to solve technical problems and translate the solution into practice (Eisenhardt et al., 2016). Thus, finding correct and effective solutions to GCs is a complex and enormously

* Corresponding author.
E-mail addresses: yelisha@pku.edu.cn (L. Ye), shan.pan@unsw.edu.au (S.L. Pan), jywang@buaa.edu.cn (J. Wang), wujj@buaa.edu.cn (J. Wu), dongxy@gsm.pku.edu.cn (X. Dong).

https://doi.org/10.1016/j.jbusres.2021.01.057
Received 23 August 2019; Received in revised form 26 January 2021; Accepted 29 January 2021
Available online 28 February 2021
0148-2963/

difficult task involving multiple types of data, multiple methods and multiple domains. For those business organizations that also engage in corporate social responsibility and pursue social impact in achieving sustainable development (Pan & Zhang, 2020), in particular, it is essential to determine how to coordinate information resources and leverage BDA to solve GCs from a management perspective.

Information triangulation is an information practice that includes information seeking, assessment and sense-making with data triangulation, investigator triangulation, theory triangulation and method triangulation (Greyson, 2018). This perspective focuses on people's information triangulation activities and examines how these activities contribute to certain outcomes (Schultze & Orlikowski, 2004). Scholars derive different types of information triangulation under various contexts; however, information triangulation with BDA for solving GCs has not been explored. In this paper, we conduct a case study to examine how BDA is deployed to address dangerous goods transportation (DGT) issues as they affect human security and sustainable cities, one of the 17 sustainable development goals provided by the United Nations, and adopt the information triangulation lens to sketch information triangulation practices from the managerial view, describing how multiple types of BD, theories and domains have been used to solve GCs.

Accordingly, our research question is as follows: How do data scientists use BDA to triangulate data, methods and knowledge in solving GCs? The contribution of this study is threefold: (1) it contributes to research on GCs and discusses how BDA can be used to solve multidomain GCs from a management point of view; (2) it enriches the theory of information triangulation and proposes several information triangulation steps in BDA for solving GCs; and (3) it also contributes some practical implications for organization management when solving social problems and pursuing sustainable development. Following this introduction, Section 2 provides a review of the literature on digital sustainability for GCs, BD and BDA and the information triangulation perspective. In Section 3, we present a case study of how a group of multidomain scientists solve DGT issues. Subsequently, Section 4 describes the case, and Section 5 offers a discussion. The conclusion and contributions are discussed in Section 6.

2. Literature review

2.1. Digital sustainability for GCs

GCs are formulations of key global health and development problems that require collaborative technical and social effort from multiple stakeholders and, if solved, will have a significant societal impact (George et al., 2016). The grand challenge as a concept was first proposed by Dr. David Hilbert, one of the most influential 20th-century mathematicians, in 1900 (George et al., 2016; Hilbert, 1902). In general, GCs include climate change, water scarcity, poverty alleviation, insect-borne disease, and global hunger (Eisenhardt et al., 2016; Ferraro et al., 2015). In 2015, the United Nations (UN) adopted a set of 17 sustainable development goals (SDGs) for solving GCs, including no poverty, zero hunger, good health and well-being, quality education, gender equality, clean water and sanitation, affordable and clean energy, decent work and economic growth, industry innovation and infrastructure, reduced inequalities, sustainable cities and communities, responsible consumption and production, climate action and so on (United Nations, 2015).

Digital sustainability has been regarded as a new goal for organizations tackling crucial sustainability GCs. It refers to organizational activities that seek to advance SDGs through the creative deployment of technologies that create, use, transmit or source electronic data (Du et al., 2013; George et al., 2019; Zhang et al., 2019). For example, Zhang et al. (2019) studied BDA capability to address air pollution management for cities' sustainability. Hu et al. (2018) proposed an integrated, multidimensional, array-based data model to address climate change by working with a variety of big climate data. Yang et al. (2017) utilized cloud computing to address the challenges associated with big geospatial data. Alam et al. (2017) illustrated how the Internet of Things (IoT) can be used for smart cities or other sustainable environments. Woolf et al. (2013) proposed instructional systems and artificial intelligence (AI) technology to develop and share global digital educational resources.

The government, nonprofit organizations and academic fields are paying close attention to digital sustainability to improve social governance, create a sustainable society and increase public welfare (Pan et al., 2021; Tim et al., 2018). In addition, an increasing number of business organizations have also become conscious of committing to social responsibility, pursuing sustainable economic, environmental and societal development, and achieving both business and social value (Pan & Zhang, 2020; Zhang et al., 2019). These enterprises increase digital input by employing more R&D staff and data scientists and develop their core products or services to generate greater societal value. However, the existing body of practices and studies on GCs are limited to challenges that belong to certain specific fields, such as meteorology, geography, or environmental science. There is a lack of research on GCs that involve multiple disciplines and require collaborative and coordinated effort from various domains to solve large social problems (Duan et al., 2019; Günther et al., 2017).

2.2. BD and BDA

BDA is a significant and important technology in digital sustainability that is currently regarded as a breakthrough technological development (Günther et al., 2017). Volume, velocity and variety are three core characteristics of BD; they represent the ever-growing dataset, which is large in magnitude; the real-time or near real-time data generation, collection and analysis process; and the various structured and unstructured data sources (Gandomi & Haider, 2015; George et al., 2016; Grover et al., 2018; Larson & Chang, 2016). In both academia and business, BD can be used for developing innovative insights, products, and services (Davenport et al., 2012). These data are complex and heterogeneous and come from public or private settings, including social media, mobile transactions, user-generated content, sensor networks, and the IoT (George et al., 2016; Sivarajah et al., 2017).

Because of the above unique characteristics of BD, there are differences between BDA and traditional data analytics in terms of data storage, data processing and analysis results (George et al., 2016). In the past, data were mostly numerical from several sources, available mainly in small quantities and stored in a database with limited capacity. They were analyzed with basic statistical tools and mainly presented with descriptive results. However, BD has higher requirements for data storage and processing technologies, and thus BDA was proposed (Constantiou & Kallinikos, 2015; Dremel et al., 2018; Gandomi & Haider, 2015; Günther et al., 2017). BDA is a set of advanced analytics methods and technologies that address BD; they can gather, analyze, link and compare large data sets, identify patterns, and generate insights from BD (Boyd & Crawford, 2012; Davenport et al., 2012; Dremel et al., 2018). BDA can be used for descriptive, predictive and prescriptive analytics, that is, to reveal the current state or pattern, forecast future possibilities, and optimize/assess the prescription identified from BD (Sivarajah et al., 2017; Strang, 2017).

In general, BDA processes have five steps: (1) data access and storage for acquiring data from a diverse source and storing it in a warehouse for value generation purposes (George et al., 2016; Sivarajah et al., 2017); (2) data cleansing and mining, for extracting and cleaning large-scale unstructured data and nonnumeric data (George et al., 2016; Sivarajah et al., 2017); (3) data fusion and integration, for aggregating and integrating cleaned data for quantitative analysis (George et al., 2016); (4) data analysis and modeling, for using various algorithms and techniques to understand the intricacy of the underlying patterns in the data (George et al., 2016; Mariscal et al., 2010); and (5) data interpretation, for interpreting the discovered data patterns and making them understandable for users, including visualizing their extracted sense and knowledge, removing redundant or irrelevant patterns and translating the useful patterns for decision makers (Mariscal et al., 2010; Simonet et al., 2015; Sivarajah et al., 2017).

Academic studies show that BDA can deliver potentially immense economic and social value (Grover et al., 2018; Günther et al., 2017; Tim et al., 2018). Regarding economic value, studies show that by adopting BD, organizations gain specific guidance for day-to-day operations and strategy orientation that can drive decision-making processes and result in increases in profit, business growth and competitive advantage (Davenport et al., 2012; Günther et al., 2017; Sivarajah et al., 2017; Tyagi, 2003). Regarding social value, studies show that BD can be analyzed to enhance information transparency, increase citizen engagement in public affairs (Kim et al., 2014), public safety and security (Newell & Marabelli, 2015), and improve education, healthcare (Cazier et al., 2015; Raghupathi & Raghupathi, 2014) and other aspects of social well-being.

In general, the existing literature is from the perspective of BD and BDA techniques, including advanced analysis tools (Chen et al., 2012; Mariscal et al., 2010) and dealing with process (George et al., 2016; Günther et al., 2017; Mariscal et al., 2010), challenges (Agarwal & Dhar, 2014; Sivarajah et al., 2017) and impact effects (Dremel et al., 2018; Günther et al., 2017). However, few studies adopt a management point of view to explore how BDA is leveraged to achieve digital sustainability (Chong et al., 2018; Zhang et al., 2019). Data scientists adopt effective triangulation practices with multiple types of data, knowledge and methods to guarantee the credibility and impacts of the analysis results. The question of exactly how different sources of information are triangulated in BDA is important for research and warrants exploration.

2.3. Information triangulation theory as the theoretical lens

Information triangulation compares data across multiple perspectives, sources or methods (Greyson, 2018). It is a kind of complex and iterative information practice; that is, it is "an array of established ways for individuals to seek and use the information available in various sources" (Savolainen, 2008). The information practice lens can be used to understand information-related activities and skills focusing on manipulating information as an object. Scholars show that information triangulation is a common practice used by scientists to improve the accuracy, credibility, and validity of findings by reducing the biases of a single approach (Greyson, 2018).

Information triangulation can be divided into four types that can be applied in problem solving: data triangulation, investigator triangulation, theory triangulation and method triangulation (Ammenwerth et al., 2003; Denzin, 2017; Wijnhoven & Brinkhuis, 2015). In data triangulation, two or more independent data sources are sought to support phenomena (Ammenwerth et al., 2003; Vikström, 2013). Investigator triangulation involves investigators from different backgrounds who are involved in researching, gathering and analyzing the data together (Ammenwerth et al., 2003; Wijnhoven & Brinkhuis, 2015). The third, theory triangulation, analyzes data based on multiple perspectives, hypotheses or theories (Ammenwerth et al., 2003). In the fourth, method triangulation, various methods are applied in the study of the same phenomenon (Ammenwerth et al., 2003; Kaulio & Mariannekarlsson, 1998). There are also two types of method triangulation, namely, between-method triangulation and within-method triangulation. Between-method triangulation involves combining different approaches to test external validity (Kaulio & Mariannekarlsson, 1998); within-method triangulation involves combining approaches from the same research tradition to cross-check for internal validity (Kaulio & Mariannekarlsson, 1998).

In the BD era, individuals may face challenges seeking, assessing and making use of information due to overload (Greyson, 2018). In particular, scholars must address heterogeneous sources, analyze data with valid methods and draw an accurate and credible conclusion. Studies show that information triangulation consists of a complex and iterative process of information seeking, assessment, and sense-making with data, investigator, theory and method triangulation, and it typically results in a decision or action (Greyson, 2018; Pee et al., 2020). Combining the BDA characteristics of high volume, variety and advanced analysis algorithms, we choose information triangulation as an appropriate lens to address our research question and to explore how BDA applied using data from multiple domains is leveraged to triangulate data, methods and knowledge in solving GCs.

3. Method

We adopt the case study methodology, which is appropriate for exploratory research (Eisenhardt, 1989; Siggelkow, 2007). In this study, we try to answer the question of "how do multidomain data scientists use information triangulation in BDA to solve GCs". The case study methodology allows us to explore new topical areas and find answers to "how" questions (Pan & Tan, 2011). We chose the DGT issue as the exemplar case of multidomain GCs and explored how to reduce the harmful effect of dangerous goods transportation on human activities; DGT involves geography, transportation planning and human activity trajectories. Dangerous goods, such as gas and hazardous chemicals transported through and around cities, can potentially harm citizens. Solving the DGT issue is beneficial to sustainable cities and communities, one of the 17 SDGs provided by the UN, as it can make cities and human settlements inclusive, safe and resilient (Hashem et al., 2016). Thus, tremendous efforts have been dedicated to dealing with DGT in both academia and governments (Wang et al., 2017). The phenomenon that we focus on in this study concerns the information triangulation practices regarding how multidomain DGT issues can be solved through BDA, as these are emerging and complex occurrences that have attracted little attention from scholars in management and business. Thus, we conducted a case study to investigate such novel phenomena in depth to develop theory and find meaningful implications (Eisenhardt, 1989).

In light of our research object, a team of data scientists solving DGT problems in China was selected as our case study for two reasons. First, this choice allows us to show how BD scientists use BDA to solve GCs that are emerging and underrepresented in the existing literature. In 2016, this team of scientists developed a novel system called DGeye to identify and predict the spatiotemporal risk patterns of DGT and to determine the underlying intrinsic mechanisms, providing an innovative way of using BDA for sustainable urban planning. Second, this case study serves as a revelatory case for exploring the information triangulation practices of data scientists in BDA. The establishment of DGeye included data triangulation involving multiple types of heterogeneous data, method triangulation with regard to data mining and investigator triangulation involving scientists with different backgrounds. Thus, this setting allows us to explore and better understand information triangulation in BDA for solving GCs.

3.1. Data collection

Our primary collected data include interview data and archival data. In January 2019, we interviewed the core members of the data scientist team as well as stakeholders in our selected case. In total, 16 people were interviewed from the School of Computer Science at University A, an Urban Planning and Design Institute, a Chinese mobile operator and the Municipal Government of City B, a mega city in China, including the DGeye project leader, project manager, data providers, data analysts, data interpreters and government officials (see Appendix B for an interviewee list). The interviews were all open-ended, flexible, and exploratory in nature; they were occasionally guided by some questions related to our theoretical lens (see Appendix C for excerpts of the interview topic guides). In summary, 25 h of interviews were conducted, and each interview lasted between 45 and 90 min. All interviews were recorded and transcribed into Chinese, and two research assistants later translated them into English to ensure data accuracy.

In addition, we collected archival data from various sources, including academic papers, news, websites and published reports. The academic papers consist of conference papers, journal papers and theses published by the data scientist team. These archival data serve as supporting evidence for triangulation, especially regarding the technical details of applying BDA in problem solving.

3.2. Data analysis

The data analysis process was inductive and iterative. We began analyzing data as we collected them to capitalize on the flexibility of the case study methodology (Eisenhardt, 1989; Pan & Tan, 2011). We conducted three rounds of data analysis. In the first round, we adopted the BDA perspective as our "sensitizing device" (Klein & Myers, 1999). The initial round of data analysis involved identifying the main stages of BDA practices in DGeye implementation. Consistent with guidelines on conducting interpretive case studies (Walsham, 1995), we performed multiple readings of the data interview transcripts, archival reports, news and academic papers to code the statements that illustrated activities related to BDA practices, such as data seeking, processing, fusion and interpretation.

In the next round, we clarified our data by summarizing the narratives about the data scientists' detailed BDA practices in tabular form and categorized the data into themes, which formed the main corpus of our subsequent analysis. Information triangulation theory provides us with the theoretical sensitivity to organize our data into themes, such as data/source triangulation and method triangulation. In addition, we derived two new themes, namely multidomain knowledge triangulation and solution-driven triangulation with the historical/current situation. We adopted existing concepts or explanations and derived new statements to gain a comprehensive understanding of the focal phenomenon. We identified the information triangulation practices for solving GCs, with statements such as identifying GC-related BD from triangulated sources, processing and stacking multi-source BD in unified dimensions, fusing BD with methodological triangulation and discovery patterns, interpreting and assessing analysis results with multi-knowledge triangulation, and solution-driven triangulation with historical/current situation.

In the third round, we tried to discover potential linkages or relationships among different information triangulation practice stages. Based on the existing literature on the BDA process and corpus, we proposed an information triangulation process in BDA for solving GCs. Except for the linear process, which runs from the beginning of BD identifying and seeking to the final application to solve GCs, we also added a stage to remove irrelevant data/information and redo the analysis between the data fusion and results interpretation stages. Finally, we organized the themes into a framework following coherent logic (Montealegre, 2002).

4. Case description

In recent years, urban safety has been an important issue for countries and regions. With the rapid agglomeration of population and industries, various types of dangerous goods, including gas and hazardous chemicals, are present in residential areas. In particular, for megalopolises such as Beijing, New York and London, a potentially catastrophic risk is posed by these goods, which can not only pollute local environments and air but also cause enormous harm to human life. For instance, on 12 August 2015, a series of explosions at a container storage station storing dangerous goods at the port of City A, a mega city in China, killed 173 people and injured hundreds of others; it was one of the largest explosions in China in recent years.

The problem of DGT has attracted great attention from the Chinese government and city planners. They care about how hazardous goods can be transported and stored away from residential areas, as well as how to predict and prevent risk occurrence. A team of scientists composed of researchers from University A and the Urban Planning and Design Institute was thus motivated to develop a computer system called "City Eyes on Dangerous Goods" (DGeye) for real-world DGT risk management (Wang et al., 2017). The researchers from University A were proficient in computer science and were mainly responsible for data seeking, data mining, system design and implementation; the researchers from the Urban Planning and Design Institute had a better understanding of geography and urban planning and were mainly in charge of specifying requirements and interpreting the results of mathematical algorithms to produce understandable explanations.

Through data mining and analysis, the DGeye results defining risk zones were consistent with historical blast sites in City A and predicted risk pattern states in City B. The scientists generated a DGT risk analysis report for the government and promoted urban reform to reduce the dangerous goods risk in cities. During the process of DGeye establishment, scientists adopted information triangulation practices by using multiple types of data, methods and knowledge and successfully deployed real-world applications. Their detailed practices are described as follows.

4.1. BD seeking and acquisition

For the scientist team, developing a system to deal with DGT was a novel and challenging issue. First, it was necessary to identify GCs related to BD. Different from the traditional top-down approaches of governments or organizations, the data scientists defined requirements in a data-driven manner by considering what types of dangerous goods-related data could be accessed and analyzed. Through preliminary discussion, two broad categories of data were identified: one category is the location of dangerous goods, and the other is the location of individuals. These two categories of data combined could be used to judge whether dangerous goods threatened human life. The project leaders talked about the significance of combining different data sources:

> *"We realize that one type of data, whether city location, DGT routes or places of human activity, can reflect only a single dimension of DGT. However, no one had put all types of data together and solved urban safety issues. Thus, we decided to consider all of them and identify whether there are conflicts between humans and dangerous goods".*

Based on these potentially accessible data, the data scientist team conducted a case study on government officers and city planners who were the users of the system to identify the needs and expectations that DGeye could address to solve GCs. The case study included three primary data sources: policies and reports, similar existing cases in other countries or cities, and interviews with core government officers and urban planners. Through case analysis, two features of DGeye could be initially determined, namely, dangerous goods risk pattern identification and prediction. The project manager, the director of the Urban Planning and Design Institute, expressed the roles of these steps:

> *"(DGT)-related policies and reports are good ways for us to grasp user needs and to specify goals in DGeye. Existing practices can provide us with inspiration on system design and feasible solutions to problems. Most importantly, we sought the opportunity to communicate directly with the mayor and secretary about our ideas and demands. They gave us their opinions, and it worked best".*

After specifying the requirements of DGeye, the data scientists decided to acquire three specific sets of data, the location data of vehicles loading dangerous goods, mobile phone signaling data of individuals and the city map, to realize risk pattern identification and prediction. The location data of vehicles loading dangerous goods are provided by the government transportation department. To track the location information of all dangerous goods, China requires that such goods be equipped with a global positioning system (GPS) terminal and report real-time locations to the local government, with this information then being aggregated to the government transportation department. The mobile phone signaling data was used to approximate the location of city residents; these data were provided by Chinese mobile operators. Mobile phone signaling data are records between mobile phones and base stations. In addition, city maps, as urban geographic data, were acquired from public sources to identify dangerous goods and the locations of individuals in the city.

In this study, the data scientists acquired two sets of data: the first was for City B from 1 January to 31 March 2015, and the other was for City A from 1 January to 28 February 2015. The project leader told us that these multiple data sets were acquired from different approaches:

> *"It is easy to obtain city maps from public channels. However, the dangerous goods transporter trajectory and mobile phone signaling data are not easily acquired. Fortunately, we had a collaboration with the government transportation department and mobile operators and could obtain these data successfully. By combining these data together, we could know the location of dangerous goods and citizens in each grid of the city moment by moment".*

4.2. BD processing and conversion

To determine when and where the dangerous goods transporter trajectory and urban citizens had conflicts, these three types of data had to be stacked in the dimensions of both time and space. Because these data are raw and heterogeneous, data processing, such as cleaning, statistical calculation and format conversion, had to be addressed before further data stacking. The data cleaning that the scientists performed mainly included identifying redundant and worthless noisy data and then replacing, modifying, or deleting them. They also had to combine real DGT situations and judge whether the data made sense. The data analysts who were responsible for DGeye establishment gave an example regarding the data cleaning:

> *"Take the dangerous goods transporter trajectory as an example. There is lots of noise in it because the data that we obtain are not a real 'trajectory', they are just a series of location points at different times. Some noisy data show that the transporter is in one place at one point but, in the next second, is in another place hundreds of miles away. This is illogical, and we have to do some mathematics to deal with it".*

After cleaning, three types of data were prepared for stacking through statistical calculation and format conversion. The city map served as the geographical base and was converted into a three-dimensional partition space, that is, longitude, latitude and time. Mobile phone signaling data were calculated to determine the population size in a specific space and time, and the data could then be stacked onto the converted city map. Dangerous goods transporter trajectory data could also be calculated to quantify dangerous goods in a specific space and time. Considering the different magnitudes of these two types of data, the data scientists calculated their respective weights. At this point, the crowd and dangerous goods location data were all stacked on the city map and could be further fused to explore risk patterns.

4.3. BD fusion and mining

Next, the data scientists fused the multiple types of data to mine risk patterns, which are defined as a set of adjacent zones that are frequently together in a risk state in the same period of time. In DGeye, two steps are performed. The first is risk pattern mining, which compresses a group of risk zones that are spatially adjacent and temporally concurrent into a relatively stable pattern. The second is causal network building, which involves delving into causal relations for risk attribution, and it can predict the state of dangerous goods risk patterns from previous states.

During this process, multiple data points were matched, calculated and integrated with algorithms and models. In risk pattern mining, latent Dirichlet allocation topic modeling, an a priori-like algorithm, and Gibbs sampling were included to generate new variables and judgments. In causal network building, the expectation maximization algorithm, priori model, likelihood model, naive Bayesian network, logistic regression, support vector machine and artificial neural network were included to triangulate which worked the most effectively in pattern state prediction. The data analysts selected a data fusion algorithm or model based on questions and their domain knowledge; the project leader said the following:

> *"Data fusion has standard methods. It's worth mentioning that when applying it to a specific question, you have to make some changes. For example, we used probabilistic graphical modeling in our previous epidemiological transmission research, but this method is not suitable for solving this DGT problem. Thus, we cannot transfer existing models into use directly but have to make adjustments according to our research setting and realistic constraining conditions".*

In total, there was a substantial amount of data being fused (City B: the mobile phone signaling data exceeded 100 G, and the dangerous goods transporter trajectory data were approximately 25 G; City A: the mobile phone signaling data also exceeded 100 G, and the dangerous goods transporter trajectory data were approximately 18 G). The above data mining algorithms and methods were able to process this BD at ever-increasing speeds and to produce DGT patterns and insights that had not been previously discovered. According to the evaluation of risky precision, the expectation maximization algorithm performed best among all algorithms and models, followed by the naive Bayesian network and the priori model.

Through data mining with the expectation maximization algorithm, some hidden patterns and their causal relationships were discovered. Fig. 1 (see Appendix A) shows the distributions of the risk patterns in City A and City B analyzed by DGeye; different colors indicate patterns of different sizes (Wang et al., 2017). Other statistics, such as the temporal distributions of risk zone proportions and the temporal distributions of risk patterns, were also calculated. In addition, expectation maximization was verified to achieve the best prediction performance compared with other models.

4.4. Result interpretation and assessment

Although the data analysts conducted data fusion and data mining in DGeye, at this point, all analysis results were in mathematical form or in the form of charts. It was difficult for government officials or city planners who were not experts in statistics to understand the results (Tham et al., 2008). Thus, these DGeye users could hardly provide suggestions or assess the analysis results from the perspective of urban safety management.

Researchers from the Urban Planning and Design Institute played an important role here. Unlike their counterparts from University A, who were proficient in computer science and mathematical analysis, the researchers from the Urban Planning and Design Institute had strong domain knowledge with regard to urban planning and could translate the mathematical results into findings understandable for government managers. The project manager emphasized their role as a data interpreter in translating between the two domains:

> *"You cannot count on all mayors or other DGeye users having a PhD and understanding statistics. You have to make sure that things are presented in an intuitive and readily comprehensive way. Thus, this is what we do here, translating realistic problems into mathematical problems for data analysts and, vice versa, translating mathematical results into understandable findings so that other domain experts can give feedback".*

In addition, the data scientists invited peers to review their results.

Regardless of the data process, data fusion or result interpretation, they provided their opinions in a modified form, which led to an iterative process to improve DGeye. Taking the risk pattern distributions in City B as an example, the director of the Urban Planning and Design Institute told us what these data interpreters did:

> *"When we get the first draft chart (of the risk pattern distribution in City B) from the data analysts, we see that there are some risk zones in certain places other than some places we take for granted. We must interpret this result by combining our knowledge of land use, infrastructure and city transportation with that of other peers. If it is hard to translate the result into a reasonable explanation, we will seek the reason in previous steps. Most often, we may even redo it completely".*

4.5. The application of DGeye

Through repeated assessment and iteration, DGeye could be successfully employed for various real-world applications. In the application to City A, DGeye showed that the dangerous goods depots were too close to residential areas, but the correlation between the number of patterns and the life rhythm of residents was very weak. The data scientists concluded that the government could monitor only some particular areas, such as the port storing chemical materials, but it could do so for a whole day. To verify its results, DGeye also accurately captured one zone as having a top-ranked risk pattern; this zone was the same as the blast site in 2015. After that incident, the government of City A took actions to change the dangerous goods depots, and it enhanced the monitoring in the depots.

In the application to City B, DGeye showed that the top-ranked risk pattern zones were located in two famous entertainment districts, and as a result of these zones, 5% of all downtown areas could be at risk. It was also found that a restaurant street in one of the districts was used for the transportation of liquefied gas cylinders. In addition, the temporal distribution of risk patterns had a rhythm similar to that of people's everyday lives. Thus, the data scientists concluded that the government should pay more attention to areas downtown in the middle of the day when people have frequent activities. The application of DGeye to City B was reported to the local government and drove the government to lay down gas pipelines for the restaurant street in 2016.

5. Discussion

In this section, we illustrate the BDA practices of the data scientists from the information triangulation perspective, we show which information triangulation practices taken by the different data scientists are combined in the BDA process, and we reveal how information triangulation involving multiple types of data, multiple methods and multiple domains in BDA can contribute to solving GCs. Fig. 2 (see Appendix A) summarizes the whole information triangulation process in BDA for GCs.

5.1. Identify and seek GC-related BD from triangulated sources

Based on our case study, we find that identifying and seeking all-round GC-related BD from triangulated sources is the first and foremost step taken by data scientists in a data-driven approach. Project leaders in the scientist team first turn the GC issues, such as the DGT issue in our case, into multidimensional data-based specific issues and identify the types of GC-related BD that are concerned with this specific problem; that is, a problem that can be largely solved by processing identified types of BD. In addition, we found that project leaders took the types of BD they already possessed or potentially could access into paramount consideration. Unlike the top-down approach, in which data requirements come from the government or the organization, this is a data-driven approach because these data provide a boundary of or constraint on the direction of data analysis, which is beneficial for the feasibility of solutions. For example, in this study, the location data of dangerous goods, urban citizens and cities that the data scientists could acquire set boundaries for further possible analysis of DGT.

Subsequently, our analysis shows that these data boundaries pave the way for specifying requirements and solutions with stakeholders for solving GCs. According to existing research, multisource requirements and related supporting evidence, such as interviews, policies and reports, from stakeholders are beneficial triangulated sources for specifying requirements and solutions (Ammenwerth et al., 2003; Vikström, 2013). In our case study, DGT requirements, such as identifying and predicting the risk patterns in the DGeye system, were specified through discussions among scientists, government and city planners and relevant secondary data. According to the specific requirements, the specified BD could then be confirmed and acquired by project managers through multiple accessible channels. To present dynamic and real-world analysis results, a longer time duration, a larger data size and smaller data granularity are preferred.

5.2. Process and stack multisource BD in unified dimensions

Processing the multiple heterogeneous data, cleaning the noisy data and obtaining well-prepared multisource BD stacks for further fusion is the second step, and the practical setting of the GC must be considered and related to the data processing. Our findings show that the data analysts responsible for data cleaning, database management and numerical calculation play an important role in this step. First, they have to clean noisy data into a standard and complete format, including deleting redundant data, modifying nonstandard data and supplementing incomplete data from various sources, which is in line with existing BDA studies (Mariscal et al., 2010; Sivarajah et al., 2017). More importantly, they have to judge the data rationality considering the real situation or context in GCs. For example, the data analysts in our case study combined real DGT situations, obtained accurate readings from the data with regard to what may happen in the real world, and judged whether the data are normal and made sense.

Next, our case study illustrates how data analysts select unified dimensions that all sources of cleaned BD can be converted into or linked to. We find that these selection criteria should be capable of supporting the realization of requirements or the goals of the solutions to GCs. For instance, to identify and predict risk zones in DGeye, the data analysts choose the spatial and time dimensions, as these can reveal the space and time conflict between DGT and urban citizens. Based on the unified dimensions, multisource BD can be calculated and stacked successfully (Sivarajah et al., 2017). Unified dimensions with various stacked data allow consideration of different aspects of one GC issue, which can provide more abundant information in further data fusion.

5.3. Fuse BD with methodological triangulation and discover information patterns behind GCs

Data fusion with methodological triangulation and the discovery of information patterns behind GCs is the third step of the information triangulation process; in this step, it is important to select and adjust relevant algorithms to make the results more applicable. According to our case study, data analysts are mainly responsible for this step, including setting up data analysis directions, selecting data fusion methods and data mining with multiple algorithms. The data analysis directions are based on GC problems according to the different requirements of solving specific GCs, such as two directions of identification and prediction in DGeye. For each direction, we have seen data analysts turn the requirements from the real world into mathematical problems, which they then refined into smaller solution steps. According to corresponding mathematical problems, standard applicable data fusion methods or improved methods can be applied to answer specific questions.

Consistent with existing studies on data fusion, various sequential or simultaneous combinations of algorithms or methods are used in multidata fusion, which is a kind of methodological triangulation (Kaulio & Mariannekarlsson, 1998). Data fusion paves the way for information mining. Some hidden patterns and causal relationships can be discovered according to the requirement of solving GCs, and they form the preliminary analysis results. For example, after DGT-related data in City A and City B are applied to data processing and mining, data analysts perform preliminary risk pattern zone identification and prediction.

5.4. Interpret and assess the analysis results with multi-knowledge triangulation

The fourth step of the information triangulation process is interpreting the analysis results and assessing them with multidomain knowledge. In this step, data interpreters serve as a vital bridge and a translator among multi-domain experts, and this part of the process may lead into iteration when there is no consensus. Existing literature shows that data interpreters who know multiple domains but are not professional analysts are important for translating different types of knowledge among experts and, in particular, for turning mathematical analysis into specific-context insights (Sivarajah et al., 2017). In our case study, the preliminary analysis results are mainly in the form of mathematical statistics or visualized charts; thus, it may be very hard for others who are proficient only in specific domain knowledge of GCs, such as experts in city planning, city safety or geography who, in our case, may have difficulty understanding statistics. Thus, it is necessary for data interpreters to translate mathematical analysis into specific-context insights.

Our analysis shows that data interpretation makes it possible for experts from different fields to understand analysis results (Mariscal et al., 2010). Through interpretation, multidomain knowledge barriers are broken, and all experts on the team can easily understand the meaning of what has been done in previous steps. These previous steps, involving data cleaning, data processing, data fusion and data mining and analysis results, can be judged and assessed by multidomain experts. If steps had errors, were not precise or the results deviated from common sense and could not be explained with specialized knowledge, it may also lead to an iteration or even a redo. Therefore, this process represents a complex but comprehensive multi-knowledge triangulation method for checking current progress and results.

5.5. Solution-driven triangulation with the historical/current situation and application to solve GCs

Finally, the information triangulation process is completed by solution-driven triangulation with the historical/current situation, wherein triangulation with the latter can generate social impact through application to solve GCs. We identify two verifications that the triangulation can complete. One is to verify the triangulated solution with historical incident-based data, which contributes to a specific GC knowledge base. Because historical incidents have already happened and are unchangeable, analyzing historical data is an effective approach to verify whether the solutions to GCs are credible. For example, in our case, the analysis results for City A coincide with a past blast event, proving that DGeye is an effective and accurate solution for the DGT issue. This is an incident-based triangulation that can provide management implications, such as how to monitor dangerous zones and prevent similar incidents from occurring.

The other approach is to verify the triangulated solution with current local-based requirements, which in turn generates policy implications. Because the current local-based condition is an everyday situation, analyzing daily data and verifying with general requirements are effective and credible triangulated approaches. This approach can also provide support for political officials or other stakeholders of GCs and generate social impact because the patterns or predictions in the analysis results can reveal problems in the current situation and drive the adoption of improved practices, such as the reconstruction of gas pipeline zones in City B by municipal government.

6. Conclusion

Our study set out to address the following research question: How do data scientists use BDA to triangulate data, methods and knowledge in solving GCs? In this study, we treat BDA as a powerful tool for digital sustainability that can solve social problems, uncover the information triangulation process and identify the crucial roles of data scientists when using BDA for GC solutions. Through a case study of how DGT issues are addressed by data scientists in China, we discover project leaders, data cleaners, data analysts, data interpreters and other data scientists playing important roles in the information triangulation process, which include identifying and seeking GC-related BD from triangulated sources, processing and stacking multi-source BD in unified dimensions, fusing BD with methodological triangulation and discovery patterns, interpreting and assessing analysis results through multi-knowledge triangulation (removing irrelevant data/information and repeating the analysis) and solution-driven triangulation with the historical/current situation and through applications to solve GCs. Our findings make several contributions.

First, we contribute to research on GCs and discuss how BDA can be used in solving multidomain GCs from a management point of view. Existing studies on GCs are dispersed across individual disciplines, such as meteorology, geography, or medicine, and most of them focus on technological details and methods (Chen et al., 2012; Eisenhardt et al., 2016; Ferraro et al., 2015; Mariscal et al., 2010). Our study adopts a management perspective and focuses on how BD can be leveraged by data scientists to form solutions for solving multidomain GCs, such as how GC-relevant BD can be identified and sought, how different algorithms and models can be triangulated to discover patterns, how multidisciplinary scientists can cooperate with domain experts from other fields, and how the ultimate solution can be verified and applied to solve GCs. Beyond technical issues, this study provides management and societal implications for data scientists who focus on GCs using BDA.

Second, we enrich the theory of information triangulation and propose detailed triangulation practices for solving GCs. Existing studies have proposed basic types of information practices, such as the seeking, assessing and sense-making of data; investigator, theory and method triangulation; and their derivations in different contexts (Ammenwerth et al., 2003; Denzin, 2017; Greyson, 2018; Wijnhoven & Brinkhuis, 2015). Our study shows a novel information triangulation practice in the context of BDA for GCs, including data source triangulation, methodological triangulation, multi-knowledge triangulation and solution-driven triangulation. It is worth mentioning that we propose solution-driven triangulation here and point out two ways of verifying triangulated solutions, namely, with historical incident-based and current local-based situations, which enrich the theory of information triangulation in the BDA context.

Third, we also offer some practical implications for business organizations aiming to solve social problems and pursue sustainable development. Many business enterprises have been conscious of committing to corporate social responsibility and trying to generate social impact beyond business profit. With an increasing number of R&D staff and scientists working in industry and finding GC solutions, it is essential to determine how to leverage these knowledge resources to solve GCs from a management perspective (Pee et al., 2010). Our study finds that some key roles are prominent in this process, for example, project leader, data cleaner, data analyst, and data interpreter. These roles are responsible for different parts of the information triangulation process. One particular role is data interpreter, who constitute a key bridge between multiple fields and are good at translating different types of professional knowledge into forms of expression that can be understood by other experts (Pee & Chua, 2016). Enterprises should pay attention to the arrangement and allocation of these core GC project

(a) Downtown Area of City B (b) Main Urban Area of City A (c) Port Area of City A

Fig. 1. Distributions of risk patterns in City A and City B from DGeye.

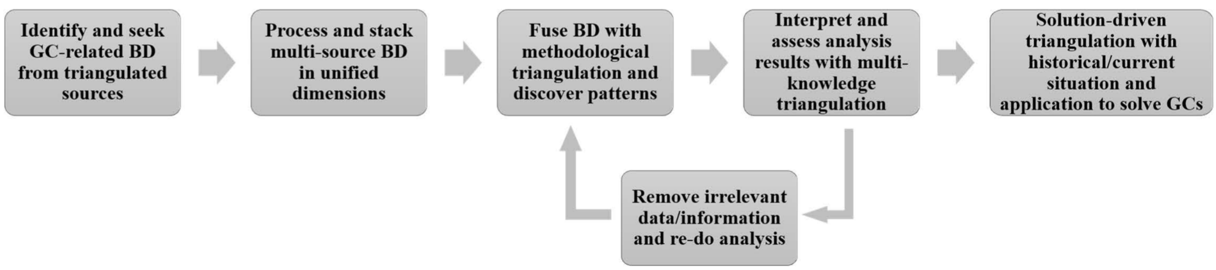

Fig. 2. Information triangulation process in BDA for solving GCs.

members.

Last, but not least, there are still some limitations in our study. First, we study the DGT issue as a case of GC and conduct a single case study. Thus, we provide the information triangulation process in BDA for solving GCs only on the basis of the DGT issue. In the future, researchers can discover other GC issues leveraging BDA and modify our findings. Second, although our study identifies some important roles in the information triangulation process, these roles play a part in relatively independent stages. It is still not clear whether and how these roles collaborate. In the future, researchers can explore the collaboration among knowledge resources in enterprises for solving GCs, which can make a further contribution to business organizations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Funding for this research is provided by National Key R&D Program of China (2019YFB2101804), National Natural Science Foundation of China (Grant Number: 71402187, 71402186) and the China Scholarship Council (CSC) (Grant Number: 201806010184). Dr. Jingyuan Wang's work was partially supported by the National Natural Science Foundation of China (Grant No.92046010, 61572059), the Fundamental Research Funds for the Central Universities (Grant No. YWF-20-BJ-J-839) and CCF-DiDi Gaia Collaborative Research Funds for Young Scholars. Dr. Junjie Wu's work was partially supported by the National Natural Science Foundation of China (71725002, 72031001, 72021001).

Appendix A. Figures

See Figs. 1 and 2.

Appendix B. List of interviewees

Interviewees	Description	Number of informants
Project leader	Professor in the School of Computer Science at University A; led the DGeye project and related urban computing projects	1
Project manager	Director of the Technology Innovation Center at the Urban Planning and Design Institute; provided project requirements and translated the data mining results for urban planning issues	1
Data providers	Director and staff of the government transportation department; collaborated with team from University A; responsible for managing GPS and dangerous goods location data	3
Data providers	Staff of Chinese mobile operators; collaborated with team from University A; responsible for managing mobile phone signaling data	2
Data analysts	Researchers in the School of Computer Science at University A; responsible for DGeye system establishment and pattern mining	2
Database administrator	Researcher in the School of Computer Science at University A; responsible for dangerous goods transporter trajectory management	1
Database analysts		2

(continued on next page)

(*continued*)

Interviewees	Description	Number of informants
	Researchers in the School of Computer Science at University A; responsible for data cleaning, data processing of dangerous goods transporter trajectory in the database	
Data interpreters	Staff members of the Technology Innovation Center at the Urban Planning and Design Institute; responsible for geographic information systems and smart city-related work	2
Government manager	Government officers of Municipal Government in City B; responsible for urban planning and construction and DGeye results assessment	2

Appendix C. Excerpts of the interview topic guides

General questions for the researchers from University A

1. Please tell us about your background and what you were responsible for in the DGeye project.
2. Please tell us about your motivation in the DGeye project.
3. What was DGeye used for? What features did DGeye provide?
4. Over how many stages was DGeye developed from an idea to a system? Which kind of work had to be done in each stage?
5. How many data sources did DGeye require? How did your team acquire the multiple types of data?
6. How were BDA methods used in establishing DGeye (e.g., multidata fusion, knowledge modeling, pattern identification and prediction)?
7. How did DGeye work in real-city application? What social impacts did it have?

General questions for the researchers from the Urban Planning and Design Institute

1. Please tell us about your background and what you were responsible for in the DGeye project.
2. Please tell us about your motivation in the DGeye project.
3. How did you collaborate with the researchers from University A?
4. What was the difference in focus between the researchers from the Urban Planning and Design Institute and those from University A?
5. How did you plan the DGeye project according to different stages?
6. What challenges did you face in the DGeye project, particularly with regard to acquiring the multiple types of data?
7. What did you do with the BDA results? How did you make them more understandable?

References

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research, 25*(3), 443–448. https://doi.org/10.1287/isre.2014.0546.

Alam, F., Mehmood, R., Katib, I., Albogami, N. N., & Albeshri, A. (2017). Data fusion and IoT for smart ubiquitous environments: A survey. *IEEE Access, 5*, 9533–9554. https://doi.org/10.1109/ACCESS.2017.2697839.

Ammenwerth, E., Iller, C., & Mansmann, U. (2003). Can evaluation studies benefit from triangulation? A case study. *International Journal of Medical Informatics, 70*(2–3), 237–248. https://doi.org/10.1016/S1386-5056(03)00059-5.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878.

Cazier, J., Cech, T., & Spaulding, T. (2015). Applying business analytic methods to improve organizational performance in the public school system. Paper presented at the applying analytic methods to improve public school performance.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Mis Quarterly, 36*(4). https://doi.org/10.2307/41703503.

Chong, M., Habib, A., Evangelopoulos, N., & Park, H. W. (2018). Dynamic capabilities of a smart city: An innovative approach to discovering urban problems and solutions. *Government Information Quarterly, 35*(4), 682–692. https://doi.org/10.1016/j.giq.2018.07.005.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology, 30*(1), 44–57. https://doi.org/10.1057/jit.2014.17.

Davenport, T. H., Barth, P., & Bean, R. (2012). How'big data'is different. https://sloanreview.mit.edu/article/how-big-data-is-different/.

Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods.* Routledge.

Dremel, C., Herterich, M. M., Wulf, J., & Vom Brocke, J. (2018). Actualizing big data analytics affordances: A revelatory case study. *Information Management and Organization Review*. https://doi.org/10.1016/j.im.2018.10.007.

Du, W., Pan, S. L., & Zuo, M. (2013). How to balance sustainability and profitability in technology organizations: An ambidextrous perspective. *IEEE Transactions on Engineering Management, 60*(2), 366–385. https://doi.org/10.1109/TEM.2012.2206113.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550. https://doi.org/10.2307/258557.

Eisenhardt, K. M., Graebner, M. E., & Sonenshein, S. (2016). Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal, 59*(4), 1113–1123. https://doi.org/10.5465/amj.2016.4004.

Ferraro, F., Etzion, D., & Gehman, J. (2015). Tackling grand challenges pragmatically: Robust action revisited. *Organization Studies, 36*(3), 363–390. https://doi.org/10.1177/0170840614563742.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. *Academy of Management Journal, 59*(6), 1880–1895. https://doi.org/10.5465/amj.2016.4007.

George, G., Merrill, R. K., & Schillebeeckx, S. J. D. (2019). Digital sustainability and entrepreneurship: How digital innovations are helping tackle climate change and sustainable development. *Entrepreneurship Theory Practice, 00*, 1–28. https://doi.org/10.1177/1042258719899425.

Greyson, D. (2018). Information triangulation: A complex and agentic everyday information practice. *Journal of the Association for Information Science Technology, 69*(7), 869–878. https://doi.org/10.1002/asi.24012.

Grover, V., Chiang, R. H., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems, 35*(2), 388–423. https://doi.org/10.1080/07421222.2018.1451951.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., … Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management, 36*(5), 748–758. https://doi.org/10.1016/j.ijinfomgt.2016.05.002.

Hilbert, D. (1902). Mathematical problems. *Bulletin of the American Mathematical Society, 8*(10), 437–479. https://doi.org/10.1007/978-1-4613-1403-5_9.

Hu, F., Yang, C., Schnase, J. L., Duffy, D. Q., Xu, M., Bowen, M. K., … Song, W. (2018). ClimateSpark: An in-memory distributed computing framework for big climate data analytics. *Computers Geosciences, 115*, 154–166. https://doi.org/10.1016/j.cageo.2018.03.011.

Kaulio, M., & Mariannekarlsson, I. C. (1998). Triangulation strategies in user requirements investigations: A case study on the development of an IT-mediated service. *Behaviour Information Technology & People, 17*(2), 103–112. https://doi.org/10.1080/014492998119599.

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM, 57*(3), 78–85. https://doi.org/10.1145/2500873.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *Mis Quarterly, 23*(1), 67–93. https://doi.org/10.2307/249410.

Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management, 36*(5), 700–710. https://doi.org/10.1016/j.ijinfomgt.2016.04.013.

Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review, 25*(2), 137–166. https://doi.org/10.1017/S0269888910000032.

Montealegre, R. (2002). A process model of capability development: Lessons from the electronic commerce strategy at Bolsa de Valores de Guayaquil. *Organization Science, 13*(5), 514–531. https://doi.org/10.1287/orsc.13.5.514.7808.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems, 24*(1), 3–14. https://doi.org/10.1016/j.jsis.2015.02.001.

Pan, S. L., Li, M., Pee, L. G., & Sandeep, M. S. (2021). Sustainability design principles for a wildlife management analytics system: An action design research. *European Journal of Information Systems, In-Press.*. https://doi.org/10.1080/0960085X.2020.1811786.

Pan, S. L., & Tan, B. (2011). Demystifying case research: A structured–pragmatic–situational (SPS) approach to conducting case studies. *Information & Organization, 21*(3), 161–176. https://doi.org/10.1016/j.infoandorg.2011.07.001.

Pan, S. L., & Zhang, S. (2020). From fighting COVID-19 pandemic to tackling sustainable development goals: An opportunity for responsible information systems research. *International Journal of Information Management, 102196*. https://doi.org/10.1016/j.ijinfomgt.2020.102196.

Pee, L. G., Pan, S. L., Li, M., & Jia, S. (2020). Social informatics of information value cocreation: A case study of xiaomi's online user community. *Journal of the Association for Information Science Technology, 71*(4), 409–422. https://doi.org/10.1002/asi.24252.

Pee, L. G., & Chua, A. Y. K. (2016). Duration, frequency, and intensity of knowledge contribution: Differential effects of job characteristics. *Information & Management, 53*(4), 435–446.

Pee, L. G., Kankanhalli, L. L. Ong, and M. K. Vu (2010) Antecedents and impact of knowledge management capability in public organizations, 14th Pacific Asia Conference on Information Systems, Jul 9-12, Taipei, Taiwan.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science Systems, 2*(1), 3. https://doi.org/10.1186/2047-2501-2-3.

Savolainen, R. (2008). *Everyday information practices: A social phenomenological perspective*. Scarecrow Press.

Schultze, U., & Orlikowski, W. J. (2004). A practice perspective on technology-mediated network relations: The use of Internet-based self-serve technologies. *Information Systems Research, 15*(1), 87–106. https://doi.org/10.1287/isre.1030.0016.

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal, 50*(1), 20–24. https://doi.org/10.2307/20159838.

Simonet, A., Fedak, G., & Ripeanu, M. (2015). Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures. *Future Generation Computer Systems, 53*, 25–42. https://doi.org/10.1016/j.future.2015.05.015.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001.

Strang, K. D. (2017). Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes? *Education Information Technologies, 22*, 917–937. https://doi.org/10.1007/s10639-016-9464-2.

Tham, Z. G., Pee, L. G., Kankanhalli, A., & Tan, G. W. (2008) Turnover in information systems development projects - managing forgetting, 12th Pacific Asia Conference on Information Systems, Jul 3-6, Suzhou, China.

Tim, Y., Hallikainen, P., Pan, S. L., & Tamm, T. (2018). Actualizing business analytics for organizational transformation: a case study of rovio entertainment. *European Journal of Operational Research, 281*(3), 642–655. https://doi.org/10.1016/j.ejor.2018.11.074.

Tim, Y., Pan, S. L., Bahri, S., & Fauzi, A. (2018). Digitally enabled affordances for community-driven environmental movement in rural Malaysia. *Information Systems Journal, 28*(1), 48–75. https://doi.org/10.1111/isj.12140.

Tyagi, S. (2003). Using data analytics for greater profits. *Journal of business strategy, 24*(3), 12–14. https://doi.org/10.1108/02756660310734938.

United Nations. (2015). Sustainable Development Goals. https://sustainabledevelopment.un.org/sdgs.

Vikström, L. (2013). Identifying dissonant and complementary data on women through the triangulation of historical sources. In *Feminism Counts* (pp. 31–42). Routledge.

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems, 4*(2), 74–81. https://doi.org/10.1057/ejis.1995.9.

Wang, J., Chen, C., Wu, J., & Xiong, Z. (2017). No longer sleeping with a bomb: A duet system for protecting urban safety from dangerous goods. *Paper presented at the proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*.

Wijnhoven, F., & Brinkhuis, M. (2015). Internet information triangulation: Design theory and prototype evaluation. *Journal of the Association for Information Science Technology, 66*(4), 684–701. https://doi.org/10.1002/asi.23203.

Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. *AI magazine, 34*(4), 66. https://doi.org/10.1609/aimag.v34i4.2490.

Yang, C., Yu, M., Hu, F., Jiang, Y., & Li, Y. (2017). Utilizing cloud computing to address big geospatial data challenges. *Computers, Environment Urban Systems, 61*, 120–128. https://doi.org/10.1016/j.compenvurbsys.2016.10.010.

Zhang, D., Pan, S. L., Yu, J., & Liu, W. (2019). Orchestrating big data analytics capability for sustainability: A study of air pollution management in China. *Information & Management, 103231*. https://doi.org/10.1016/j.im.2019.103231.

Lisha Ye is a researcher in the Department of Strategy and Industry, China Mobile Research Institute. She received her PhD in Management Science from Peking University in China. Her research interests are emerging digital enablement phenomena in businesses and societies. She has conducted in-depth case studies on big data analytics, social innovation, start-up companies and large commercial organizations. She has published articles in peer-refereed journals such as *International Journal of Information Management* and some leading management field conferences.

Shan L Pan is AGSM Scholar and a professor of Information Systems and Technology Management at the UNSW Business School. He is the founder and leader of the Digital Enablement Research Network (DERN) at UNSW Business School, a research group focused on transformational effects of technologies and consumerization effects of technologies. He is currently affiliated with the Alexander von Humboldt Institute for Internet and Society, Germany. He is also an advisor to Alibaba Research Institute, China. His research interest is digital enablement in business and social innovation. Professor Pan's conducted in-depth studies on state-owned enterprises, commercial organizations, rural villages, and non-profit organizations in China, Germany, Finland, India and Southeast Asia. He has published widely in top journals such as *MIS Quarterly, Information Systems Research, Journal of the AIS,* Information *Systems Journal, European Journal of Information Systems, Journal of Strategic Information Systems, Journal of Information Technology, Journal of the Academy of Marketing Science, Journal of the Association for Information Science and Technology, European Journal of Operational Research, IEEE Transactions on Engineering Management, Communications of ACM,* and among others.

Jingyuan Wang, the head of BIGSCity (Bigdata Intelligence Group on SmartCity), is an Associate Professor in School of Computer Science and Engineering, Beihang University. He got his Ph.D. degree in 2011 from Computer Science Department, Tsinghua University. He published more than 20 papers on top journals (e.g., *IEEE TKDE, ACM TOIS, IEEE TMC, IEEE Intelligent Systems*) and conferences (e.g., *ACM KDD, IEEE ICDM, AAAI, IEEE Infocom, ACM CIKM*), as well as named inventor on several granted CN and US patents. His research interests include AI, data mining, and urban computing.

Junjie Wu is a professor in Information Systems Department, School of Economics and Management, Beihang University. He received his Ph.D. degree in Management Science and Engineering from Tsinghua University. His general area of research is data mining, with a special interest in social, urban and financial computing. He has published one monograph in Springer and over 100 papers in refereed conference proceedings (e.g., *KDD, IJCAI, AAAI*) and journals (e.g., *ISR, IEEE TKDE, IEEE TIP*). One of his papers was selected as a ten-year highly cited paper by Essential Science Indicators (ESI). He is the recipient of various nation-wide academic awards, e.g., the NSFC Distinguished Young Scholars, the MOE Changjiang Young Scholars, and the National Excellent Doctoral Dissertation.

Xiaoying Dong is an associate professor in Guanghua School of Management, Peking University. She received her PhD in Information Science from Peking University. She had academic visits at Harvard University, University of Pittsburgh, and Australian National University. Her research focuses on enterprise knowledge management, digital transformation, and CIO leadership. She has published some books and a number of articles in peer-refereed journals such as *Journal of Business Research, International Journal of Information Management*, etc.