The use of heterogeneous geospatial big data for improving decision-making

Flávio E. A. Horita[1,2]**, João Porto de Albuquerque**[1,3]

[1]Institute of Mathematical and Computing Sciences (ICMC)
University of São Paulo (USP) – São Carlos — Brazil

[2]Center for Mathematics, Computation and Cognition (CMCC)
Federal University of ABC (UFABC) – Santo André – Brazil

[3]Center for Interdisciplinary Methodologies (CIM)
University of Warwick – Coventry - U.K.

`flavio.horita@ufabc.edu.br, jporto@icmc.usp.br`

Abstract. *Emerging data sources have been increasing the amount of available and useful data, which have a potential for revolutionizing entire business processes and decisions in several scenarios, e.g., smart cities. However, at the same time that these "big data" open further opportunities, the heterogeneity of their features often hampers the integration and visualization of data. Therefore, this work presents an approach to handle heterogeneous geospatial big data for supporting a more informative decision-making. Study results advanced the state-of-the-art by understanding decision-makers' requirements and developing innovative decision support system. These indeed provided valuable contributions to practice and research.*

1. Introduction

Decisions are an intrinsic part of the daily running of organizations; managers are continuously analyzing business variables to choose the best alternative for increasing profits and/or reducing production costs. Available information plays an essential role in improving the effectiveness and accuracy of those choices, and thus new technologies and mechanisms have been developed to enhance data collection. This allows decision-makers in understanding, and indeed knowing, their business, which may lead to improved decision-making.

All these "big data" can be characterized by four factors [Hashem et al. 2015]: (1) a huge volume of data; (2) traditional and regular databases that are not able to handle the available data; (3) the rapid growth in the volume of data; and (4) the concern in regarding the veracity of the available data. **Related works** indeed investigated the question of integrating these data has been widely investigated in the literature [Hashem et al. 2015, Wamba et al. 2015]. [Schnebele et al. 2014] describe distinct approaches for integrating authoritative and non-authoritative data to provide location-based eventful visualization, a statistical analysis and graphic capabilities for the authorities and the public.

Big data definitely open further opportunities to support daily activities of organizations, however at the same time they also raised some challenges. Here, the **challenges** we seek to examine are twofold. First, the huge volume of geospatial data, together with

a heterogeneity of formats, often hampers their appropriate integration and suitable visualization [Vieweg et al. 2014, Schnebele et al. 2014]. This leads to data being overlooked by managers and directors and become useless to support decision-making. A decision support system should be designed for dealing with all idiosyncrasies of heterogeneous geospatial big data sources. Second, data can be of great value in supporting decision-making only when they reach the decision-makers in a suitable way; otherwise, they might become useless or require extensive knowledge and experience for further data processing. Understanding decision-makers' information needs and their connection with data sources is not only valuable but also essential for supporting decisions that use geospatial big data. In light of these challenges, this research sought to answer the following **research question**: *how can the integration[1] of heterogeneous data sources contribute to the improvement of decision-making?*

For addressing this question, the **objective** of this work is to establish an approach that improves decision-making with heterogeneous big data by understanding decision-makers' requirements and developing innovative decision support system. This approach was built on three supplementary elements: a) a conceptual architecture that allows integrating heterogeneous data sources; b) a model-based framework that connects decision-making with appropriate data sources; and c) a set of design principles that must be observed to ensure that a spatial decision support system can effectively support decision-making with heterogeneous geospatial big data. The context of disaster management in Brazil has been adopted as our case study in this workHence, the **motivational factors** behind this work are both academic and practical as it investigates research challenges with regard to the decision-making using heterogeneous geospatial data sources. While, on the other hand, project results may provide a technological and theoretical framework for supporting emergency agencies in the mitigation and response to disasters.

The work is organized as follows. Section 2 details the study context, while Section 3 presents the obtained results. Section 4 describes the contributions and impacts of this work. Eventually, Section 5 summarizes our conclusions and makes recommendations for future work.

2. Study Context

The **study context** of this work lies on the tasks of early warning and environmental monitoring as they are the most important tasks of preparedness phase in disaster management. While the task of early warning is responsible for analyzing the hazards, risks, and vulnerability of communities and issuing warnings to ensure that appropriate response capabilities are in place at the right time, its accuracy relies on the quality of data that are shared by available monitoring systems, as well as on the extension of monitored area. In Brazil, these tasks are undertaken by the National Center for Monitoring and Early Warning of Natural Disasters (Cemaden) in Brazil using a monitoring sensor network that includes more than 4,750 rainfall gauges, about of 550 humidity and rainfall sensors, nine weather radars, and almost 300 hydrological stations. These sensors provide data about the precipitation, calculate the motion of weather systems, and estimate the weather type (e.g., rain). Furthermore, citizens of vulnerable communities collect data from semi-automated

---

[1]Here, the term "integration" is understood as the combination of data provided by different sources, as well as the visualization of the integrated data in a simple and unified view. It does not include database theories or on data mining/data science techniques.

rainfall gauges installed in the region of these communities and share them with the center via a web-based system. Around 1,000 of these gauges have been installed in different communities in Brazil. Although Cemaden has its own monitoring facilities (including hydrological stations, meteorological stations, and automated rainfall gauges), the center also works in collaboration with several institutions, e.g., the National Water Agency (ANA, in Portuguese). These provide further data that thus supplement the existing data of the center. Members of a control room use a decision support system for monitoring environmental variables, which integrates data from monitoring systems and displays integrated data using a geospatial dashboard.

3. Results

Key results achieved here are manifold. To begin with, this work has been set out a conceptual architecture that is based on interoperability standards for integrating this volunteered information with official (and traditional) data. Lessons learned from its implementation and employing to support decision-making in a real-world scenario for flood risk management showed that volunteered information is able to supplement official data. This finding of the use of distinct data sources adds to previous work [Schnebele et al. 2014] insofar it employs interoperable standards so that data sources could be included in a more flexible way. Study results also demonstrated that these interoperability standards in combination with data source adapters are important elements of a layer-based architecture because they are able to handle intrinsic heterogeneity of the data sources. This integrated use is performed by building upon and extending previous work that uses adapters for integrating geospatial data [Assis et al. 2016]. In other words, those adapters are able to encapsulate the discrepancies and idiosyncrasies of data sources, converting the data to a standardized pattern, and sharing them. Besides, results obtained in this work also showed that a proper data integration should match target-users' requirements; otherwise, they may become useless or overlooked.

Based on results of the previous work, we designed an integrated model and notation, named oDMN$^+$, which describes a) how business tasks are connected to decisions, then b) which information is required by them, and finally c) what data source could be used for providing this required information. It was based on a standard model and notations from business management (e.g., BPMN and DMN) in combination sensor data and observations (e.g., O&M). Lessons were learned from a case study in a Brazilian early warning center for modeling the decision-making that is required for issuing warnings to vulnerable communities about impending disasters. This case study showed that oDMN$^+$ is able to provide an understanding of how to make better use of the available data sources that provide useful information for decision-making. It can also assist in locating those activities where there is a lack of information and in finding the available data sources that can help. These findings supplemented existing studies on the question of making a connection between decision-making and data sources [Kleindienst et al. 2015] by providing a notation that is able to describe data sources and connect them to the information required to make a decision. They also support and improve the use of the analytical models proposed in previous works. Study findings also pointed out that a modeling process is essential for successfully obtaining conceptual elements from the decision-makers in the context of an application, as well as enabling the oDMN$^+$ to be employed in practice. This adds to previous research works on business modeling [Santoro et al. 2010] by

proposing and evaluating a modeling process that is not only focused on modeling business processes but also of modeling variables associated with decision-making. The obtained results also suggest that decision support systems in the context of big data provide features that allow end-users to control the visualization of data sources of interest. This is particularly important since the required information and data sources often change.

Finally, a set of design principles was formulated to assist the design of a spatial decision support system which could be applied to decision-making with heterogeneous geospatial big data. These were derived from a cross-collaborative action research project that was conducted in collaboration with a Brazilian early warning center. The purpose of these principles was threefold: (a) to guide the development of spatial decision support system, (b) to embody the decision-makers' requirements within the system, (c) and to improve decision-making with heterogeneous geospatial big data. Within the scope of the work, we evaluated the design principles with members of the monitoring control room of a Brazilian early warning center. Study findings suggest that practitioners and researchers should understand how these data would be useful for fulfilling their information requirements before starting the use of geospatial big data. This plays a crucial role in avoiding the visualization of information that is not required for decision-making and thus might become useless. Although some works have been done on analyzing the use of big data in decision-making [Vieweg et al. 2014, Wamba et al. 2015], they lack in describing a design process for decision support systems that use heterogeneous big data sources. This can be overcome by the design principles developed in this work, which provide suitable guidelines and important factors to be analyzed when designing those systems.

4. Contributions and impacts

For a better means of recognizing and describing these "contributions", we relied on another distinction drawn by [Gregor and Hevner 2013] between descriptive (denoted by Ω), which represents the knowledge of natural phenomena and the laws and regular patterns of phenomena ("what"), and prescriptive (denoted by Λ), which consists of the knowledge of human-built artifacts ("how"). A Λ knowledge can be represented by the following types: constructs, models, methods or instantiations. Overall, this work provides **key contributions** to both types of knowledge as they advance the current state-of-the-art on the use of heterogeneous geospatial big data to ensure more effective decision-making.

Contributions and impacts to research. Results obtained in this work were published by means of papers and articles in conferences and journal of different domain areas, e.g., Computers & Geosciences (QUALIS-CC A1), Decision Support Systems (QUALIS-CC A1), International Journal of Disaster Risk Reduction (IF: 1,603), AMCIS (QUALIS-CC A2), and HICSS (QUALIS-CC A1). These publications thus showed the interdisciplinarity of the results, and thus contributions to the different body of knowledge. The following list presents the main publications:

- **[Horita et al. 2018] International Journal of Disaster Risk Reduction (IF: 1,603, h-index 28).** It provides an understanding of factors that influence decision-making within control room of a Brazilian early warning agency. Results have impacted the literature of disaster risk management and studies of control rooms.
- **[Horita et al. 2017] Decision Support Systems (IF: 3,222, QUALIS-CC A1, GoogleScholar=8).** This article describes oDMN$^+$ framework and its evaluation in

a case study conducted within a Brazilian early warning agency. It impacts the existing literature on decision-making and information systems.

- **[Horita et al. 2015] Computers & Geosciences (IF: 2,553, QUALIS-CC A1, GoogleScholar=53).** This work introduces the conceptual architecture and describes lessons learned from its employing in a real-world case study of flood risk management. It impacts existing literature on the integration of heterogenous data sources.
- **[Horita et al. 2013] Americas Conference on Information Systems (AMCIS) (QUALIS-CC A2, GoogleScholar=67).** It details results of a literature review that analyzes the use of volunteered information to support disaster management.

In addition, six conference papers, two workshop abstracts, and one more journal article also published results obtained in the work. Other publications are one book chapter, one journal article, and eight conference papers, which can be found in the PhD Thesis. It also received a honorable mention.

Contributions and impacts to practice. In particular, empirical evidence has shown that the conceptual architecture represents a means of integrating heterogeneous data sources, sharing the integrated data, and visualizing them within a simple Web-based DSS. Furthermore, using oDMN$^+$, decision-makers can understand the interplay between their business process and existing data sources, while the design principles support the development of DSS that supports decision-making with heterogeneous geospatial big data. It is also worthwhile to remark that the project results also provided valuable contributions to organizations, e.g., for supporting interoperability of systems in Brazilian Meteorological Agency (CLIMATEMPO). Moreover, the National Centre for Disaster Monitoring and Early-warning of Natural Disasters (Cemaden) was impacted by the results of this work as they provide a solid basis for conforming to the guiding principles that were defined in the Sendai Framework for Disaster Risk Reduction 2015-2030 [UN 2015] and for assisting the understanding of decision-making process within the control room.

5. Conclusions

This work presented an approach that comprises the following new and innovative elements: 1) the conceptual model that integrates heterogeneous data sources; 2) a model-based framework responsible for connects decision-making with data sources; and 3) design principles that supports the design of a decision support system to effectively assist decision-making with heterogeneous geospatial big data. By taking disaster management as a representative scenario of decision-making based on heterogeneous geospatial big data, two different empirical case studies were conducted for evaluating the produced elements: one in the context of flood risk management in São Carlos and another one within the control room of a Brazilian early warning center. Empirical results indeed provided evidence that the conceptual architecture determined essential software components for handling the intrinsic heterogeneity of the data sources (e.g., data source adapters). These components are enhanced by a better understanding of the target users' needs, modeled using oDMN$^+$. The development of decision support systems is improved by the design principles that can help in making a better use of heterogeneous geospatial big data.

References

Assis, L. F. F. G., Behnck, L. P., Doering, D., Freitas, E. P., Pereira, C. E., Horita, F. E. A., Ueyama, J., and Albuquerque, J. P. (2016). Dynamic sensor management: Extending

sensor web for near real-time mobile sensor integration in dynamic scenarios. In *Proceedings of the 30th International Conference on Advanced Information Networking and Applications (AINA)*, pages 303–310.

Gregor, S. and Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2):337–356.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, 47:98–115.

Horita, F. E., Albuquerque, J. P., Marchezini, V., and Mendiondo, E. M. (2017). Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in brazil. *Decision Support Systems*, 97:12–22.

Horita, F. E., de Albuquerque, J. P., and Marchezini, V. (2018). Understanding the decision-making process in disaster risk monitoring and early-warning: A case study within a control room in brazil. *International Journal of Disaster Risk Reduction*, 28:22–31.

Horita, F. E. A., Albuquerque, J. P., Degrossi, L. C., Mendiondo, E. M., and Ueyama, J. (2015). Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. *Computers & Geosciences*, 80:84–94.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. G., Zipf, A., and Albuquerque, J. P. (2013). The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: a systematic literature review. In *Proceedings of the 19th Americas Conference on Information Systems (AMCIS)*, pages 1–10.

Kleindienst, D., Pfleger, R., and Schoch, M. (2015). The business alignment of social media analytics. In *Proceedings of the 23rd European Conference on Information Systems (ECIS)*, pages 1–14.

Santoro, F. M., Borges, M. R. S., and Pino, J. A. (2010). Acquiring knowledge on business processes from stakeholders' stories. *Advanced Engineering Informatics*, 24:138–148.

Schnebele, E., Cervone, G., and Waters, N. (2014). Road assessment after flood events using non-authoritative data. *Natural Hazards and Earth System Sciences*, 14(4):1007–1015.

UN (2015). Sendai Framework for Disaster Risk Reduction 2015-2030. Technical report, United Nations. Technical Report.

Vieweg, S., Castillo, C., and Imran, M. (2014). Integrating social media communications into the rapid assessment of sudden onset disasters. In *Proceedings of the 2014 International Conference on Social Informatics (SocInfo)*, pages 444–461.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., and Gnanzou, D. (2015). How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165:234–246.