Big-Data and the Question of Horizontal and Vertical Intelligence: A Discussion on Disaster Management

Frederick Benaben[1], Aurelie Montarnal[1], Audrey Fertier[1], Sebastien Truptil[1],

[1] Toulouse University, Mines Albi – Campus Jarlard, Route de Teillet, 81000 Albi - France
{frederick.benaben, aurelier.montarnal, audrey.fertier, sebastien.truptil}@mines-albi.fr

Abstract. Due to the unexpected context of disaster management (with heterogeneous and non-dedicated potential data sources), the classical Big-Data approaches within this business domain (schematically focused on data storage and pattern recognition) appear to be limited, especially when trying to get a situational level of such crises from data gathered from the field. That is why this article aims at discussing a specific vision of Big-Data for data management, in two steps: (i) analyzing this business domain to identify relevant characteristics, impacted or concerned by Big-Data, and describe the new key challenges that need to be tackled, and (ii) designing an innovative Big-Data framework dedicated to this particular business domain. After having highlighted the importance to push abstraction levels and especially data interpretation as a way to perform vertical intelligence in data analysis (instead of horizontal intelligence with usual approaches), the proposed Big-Data framework brings a layered approach according to three dimensions: gathering (data level), interpretation (information level), exploitation (knowledge level).

Keywords: Big-Data, Collaborative Situation, Data Interpretation, Metamodel, Modeling, Intelligence, Disaster Management.

1 Introduction

Just like any domain of human civilization, Disaster Management (DM) has to take into account the unavoidable increasing of generated and potentially accessible data. On any crisis site, heterogeneous and potentially not-dedicated data sources (sensors, opendata, witnesses and social networks, etc.) generate a deluge of data. Those emerging data could be very useful to characterize crisis situations and, thus, allow to better support DM and the decision-makers awareness. However, dealing with their specificities leads to raise new key challenges, and to the necessity of implementing a new Big-Data approach adapted to the needs of this business domain. That question of Big-Data is usually considered through the angle of emerging new technological solutions and platforms to manage this unprecedented amount of data. Though, the

[1] Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

question of Big-Data definitely concerns so many human and social domains that it could be relevant to imagine that all domains do not require the same technological (or philosophical) approach for Big-Data management.

Hence, this article aims at giving a vision of Big-Data applied to the DM, by exploring those various challenges raised both in terms of technological and philosophical points of view. Answering those issues leads to discuss the necessity to implement a 'vertical intelligence' and in particular, the ways and the means to bring levels of abstraction to the gathered data. Concretely, three main questions (and levels of abstraction) are addressed: (i) data gathering - how to extract data from unknown data sources and deal with their veracity and usability? (ii) data interpretation - how to transform datasets into exploitable models? and (iii) information exploitation - how to transform crisis situational models into actionable knowledge?

According to these objectives, this article is structured according to three sections: first section describes specificities and characteristics of DM from the Big-Data perspectives in order to define a particular profile for this domain. Second section uses this domain profile to infer what specific features of Big-Data management would be required in DM context (with regards to usual ways of tackle the question of Big-Data in other domains) and to introduce a Big-Data for DM framework. Finally a conclusion discusses the usage of this vision of Big-Data.

2 Specificities of Disaster Management on Social and Technological Perspectives

Based on [1] and [2], [3] the following definition for Disaster is suggested: *disruption within the state of a system, which reveals instability and discontinuity and which requires a specific treatment to deal with the unwanted consequences and to obtain a new acceptable state of the considered system*.

2.1 Social Dimension

Considering that consequences and impacts of Disasters are usually multi-dimensional and concern as various stakes as public health, politics, economics, etc., various types of stakeholders must be involved in DM (to cover all the involved multi-dimensional treatment actions). These responders should moreover be coordinated and smoothly synchronized to avoid conflicts and frictions due to pursuing heterogeneous missions and objectives. Consequently, dedicated collaborative networks must be established to perform efficient and adequate response. However, the usual questions related to the use of collaborative networks are exacerbated, due to five reasons:

- Critical stakes (consequences of crises may be heavy as far as they concern human lives and major societal stakes).

- Fickle kinetics (crisis is a fully dynamic system with cascading effects and potentially chaotic behaviour).
- Incomplete vision (stakeholders involved in crisis management only get a partial picture of the entire situation and its consequences).
- Unpredictable aspect (even if often expected, a crisis generally occurs at an unpredicted time that requires means and resources to be ready and available).
- Unheard of aspect (crisis management is not a mastered recurrent process and stakeholders, even if prepared at a micro-level are not a macro-level).

Furthermore, crisis management, as collaborative situation, concerns heterogeneous partners, which have their expertise area and are fully competent at a micro-level (besides it is their everyday-life activity). Consequently, and as described in [3], the main issue concerns coordination and collaboration of these stakeholders at a macro level. Finally, the main characteristic of DM from a **social perspective** is that *building efficient collaborative networks to support partners' coordination, although the supreme objective, is far more difficult in DM than in other domains.* This statement also includes all cultural barriers, rivalry and other human lows exacerbated by the critical context.

2.2 Technological Dimension

Another crucial element to be considered concerns the constantly increasing volume of available data. This is an indisputable matter of fact: today's world is numeric. This *cliché* also needs to be taken into account when considering DM: stakeholders are facing an increasing number of data sources, including non-dedicated ones. Practices, tools and usages regarding data management in crisis contexts are not appropriate to this situation anymore and must be adapted. Disaster managers must now deal with the following issues:

- In DM context, in addition to the (very classical) wide range of data types (sensors measures, photo, report, social network message), and in contrary to a lot of other domains, data sources are often massively unknown while starting the job.
- Credibility and trust of (potentially unknown) data sources must be evaluated [5] and [6].

In [7] the authors particularly describes how big-data can be considered as a new challenge in crisis management, but not only in a "classical" way (mainly considering the increasing amount of data and dataflow): the way this large amount of data could be treated to become intelligible to exploit it relevantly (automatically or not) is at the

heart of this new challenge. Finally, the main characteristic of DM from a **technological perspective** is that *performing Big-Data management implies to deal with unknown and heterogeneous data sources, potentially non-dedicated ones*. In DM, discovery of data sources and understanding of data is probably as important as exploiting data to support decision and action.

2.2 Conclusion of DM Characterization

The specificities of Big-Data management in DM can finally be summarized according to the following sentence: *The question is mainly about managing unknown data sources (rather than large amount of data coming from well-known data sources) to support solely and mainly coordination issues (rather than a multi-objectives exploitation)*. It is important to understand that both these specificities are really different from classical Big-Data problems:

- Data sources are generally well known: For instance, the industrial domain (from supply-chain to production) is based on a lot of data (such as the amount of stock, the location of a deliverer, the customer's invoice, etc.) that can be easily monitored. The only unexpected data sources could be some borderline examples, such as those related to weather forecast or traffic topics.

Data exploitation is generally covering a wide range of elementary purposes: For instance, in public transportation domain, Big-Data management may be dedicated first to ensure transportation continuity through management of inter-mode solutions (offering reactively palliative solutions in case of failure of one mode), second to manage customer relationship (CRM system), third to improve the network efficiency, etc.

3 A Specific Big-data Framework for Disaster Management

Classical Big-Data problems can be summarized (from a very simple and schematic perspective) according to two main goals: it is mainly about (i) increasing the data storage capacity (to be able to collect and store as much data as possible, including historical consideration) and (ii) improving pattern recognition feature (to be able to parse the obtained gigantic data fields and find quickly and correctly specific patterns). Most technological progresses made these years in Big-Data domain are dedicated to one or both of these expectations (more storage or more parsing). Actually, this approach is perfectly adapted to most of the domains in which Big-Data is one perspective of improvement because: (i) data sources are well known and (ii) exploitation of data field through pattern recognition can serve a lot of simple objectives. However, as presented in the first section, this is exactly what DM is not.

In [7] the authors particularly describes how big-data can be considered as a new challenge in DM, but not in a "classical" way (mainly considering the increasing amount of data and dataflows): the way this large amount of data could be treated to become intelligible to exploit it relevantly (automatically or not) is at the heart of this new challenge. This article considers that Big-Data for DM should perform the following specific features.

3.1 Gathering: from Data Sources to Data Storage

Considering that data sources on disaster site might be unknown, randomly distributed and potentially doubtful in terms of veracity, the first layer of the Big-Data framework is in charge of identifying the data environment to be considered. This objective is a very innovative one and really arduous on a technological point of view; actually, Big-Data approaches are usually based on predefined data background with selected data sources.
The expectation is to perform *discovery* (of unknown data sources), *understanding* (of the emitted data), *trust and veracity evaluation* (of the understood data) and *subscription* (to discovered, understood and accepted data sources) mechanisms. It is absolutely necessary, in the connected world of DM, to make the gathering layer open, *i.e.* to allow Big-Data for DM to deal with a dynamical data environment involving incoming and outgoing unknown and non-dedicated data sources.

3.2 Interpretation: from Data to Information

The main stake for Big-Data in DM is probably the ability to interpret gathered data in order to automatically build situational models (such situational model might be considered as a common operational picture). This step is really the corner stone of Big-Data for DM. Actually, while classical Big-Data aims at exploring gigantic fields of data to detect patterns and thus at providing some direct results, DM Big-Data must insert an abstraction level between raw data and exploitation results. This intermediary level is in charge of combining and transforming collected data into intelligible concepts. This level is mandatory in the case of DM because data sources and data types may be unknown and consequently not directly exploitable.
The theory defended in this article states that this information level, to perform interpretation requires a situational metamodel (presenting concepts and relations inherent in DM domain). Such a metamodel (see the one described in [3]) provides the theoretical background to create a situation model by instantiating concepts and relations with gathered data.

3.3 Exploitation: from Information to Knowledge

Finally, the last level for Big-Data for DM is about the ability to exploit fruitfully the generated and formalized information, to support the coordination of responders. This

level concerns mainly exploitation features to perform *decisions* and *actions*. Actually, there are a lot of DM tools covering coordination functionalities (such as first-responders alert, collaborative tasking, monitoring of coordinated actions, or orchestration of collaborative workflows), which would strongly benefit from such information base: these tools require real inputs to perform their service; consequently, providing these tools with consolidated information basis would remove this constraint and furthermore would ensure a higher quality of service considering that the collected and interpreted data would certainly supply robust and trustable situation models.

3.4 Big Picture of the Presented Big-Data Framework for Disaster Management

On the following figure, the three levels presented above are conceptually presented.

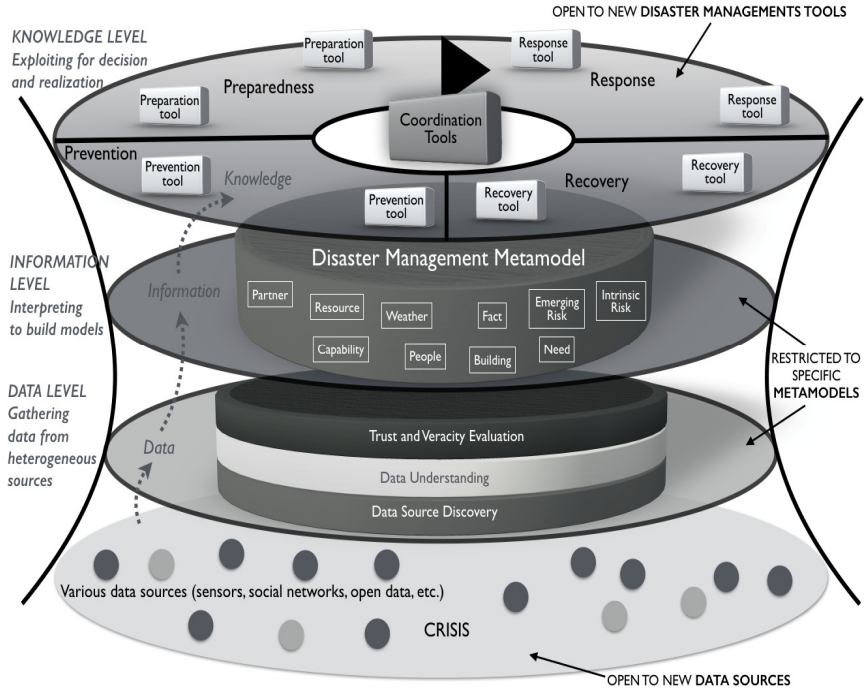


Fig. 1. Big Picture of the Big-Data for DM framework.

The most noticeable elements of figure 1 are the following:

- The abstraction path from data to information to knowledge inherent in the Big-Data for DM.
- Data level and Knowledge level are supposed to be open (allowing the use of emerging data sources and the exploitation of new DM tools).

- The central role of the Metamodel to perform data interpretation and information formalization.

4 CONCLUSION

This article aimed at presenting an original vision of Big-Data specifically dedicated to Disaster Management. Actually, this vision may be relevant for various domains as far as they match with criteria identified in first section. Big-Data is definitely a way to provide intelligence to data gathering. The main goal of big data as presented in [8] and [9] is to facilitate the decision-making, through a swift analysis of large amount of data, of different types, from a variety of sources, to produce a stream of actionable knowledge. Consequently, it is dedicated to provide a computed system with the ability to exploit efficiently and relevantly large amount of data. However, the important aspect of the vision presented in this article is that it opposes an horizontal vision of this intelligence (mainly dedicated to increase on the one hand data storage and indexation capacities, and on the other hand, parsing and digging capacities) to a vertical vision of this intelligence (dedicated to stack up abstraction levels to improve interpretation capacity). The position of this article is to respect both these visions, but to claim that Disaster Management is a very specific domain, with a lot of particular characteristics (such as the unknown aspect of the impacted perimeter or the very wide potential origin of data from sensors to social networks). Consequently, one purpose of this article was to show that vertical intelligence in Big-Data is potentially better adapted to Disaster Management.

References

1. Tomasini, R., and Van Wassenhove, L. N. (2004) Logistics response in the 2002 food crisis in Southern Africa. INSEAD Case.
2. Devlin, E. S. (2006) Crisis Management Planning and Execution. CRC Press.
3. Benaben, F., Lauras, M., Truptil, S. and Salatge, N. (2016) A Metamodel for Knowledge Management in Crisis Management, *Proceedings of 49th Hawaii International Conference on System Sciences*, pp. 126-135.
4. Camarinha-Matos, L. and Afsarmanesh, H. (2005) Collaborative networks: A new scientific discipline, *Journal of Intelligent Manufacturing*, Springer, 16 (4), pp. 439-452.
5. Truong, B., Caragea, C., Squicciarini A, and Tapia, A. H. (2014) Identifying valuable information from twitter during natural disasters, *Proceedings of the American Society for Information Science and Technology*, 51(1), pp. 1-4.
6. Tapia, A. H., Moore, K. (2014) Good Enough is Good Enough: Overcoming Disaster Response Organizations' Slow Social Media Data Adoption, *Computer Supported Cooperative Work*, 23(4), pp. 483-512.
7. Meier, P. (2013) Strengthening humanitarian information: the role of technology, Chapter 3 of World Disaster Report 2013, International Federation of Red Cross, Red Crescent Societies (ICRC).
8. Kalyvas, J. R. (2015) A Big Data Primer for Executives, *In Big Data*, CRC Press.

9. Power D. J. (2014) Using ‘Big Data’ for analytics and decision support, *Journal of Decision Systems*, 23(2), pp. 222-228.