Adoption of big data in crisis management toward a better support in decision-making

Audrey Fertier
Centre Génie Industriel, Université de Toulouse - Mines Albi, Campus Jarlard, 81000 Albi, France
audrey.fertier@mines-albi.fr

Aurélie Montarnal
Centre Génie Industriel, Université de Toulouse - Mines Albi, Campus Jarlard, 81000 Albi, France
aurelie.montarnal@mines-albi.fr

Anne-Marie Barthe-Delanoë
Université de Toulouse, INP-ENSIACET & CNRS, LGC (Laboratoire de Génie Chimique),
4 allée Emile Monso
31432 Toulouse, France
annemarie.barthe@ensiacet.fr

Sébastien Truptil
Centre Génie Industriel, Université de Toulouse - Mines Albi, Campus Jarlard, 81000 Albi, France
sebastien.truptil@mines-albi.fr

Frédérick Bénaben
Centre Génie Industriel, Université de Toulouse - Mines Albi, Campus Jarlard, 81000 Albi, France
frederick.benaben@mines-abli.fr

ABSTRACT

Most agree that the innate complexity and uncertainty of a crisis compel the stakeholders to coordinate in a hurry, despite their heterogeneity or the volume of data to process. Supporting their coordination is now possible, thanks to a mediation system combined with big data management tools. The GéNéPi[1] project explores this possibility and proposes to improve the generation of collaborative processes offered by the MISE[2]'s solution. The idea is to increase the number of usable data sources. To do that, in a fixed time-frame, the situation models have to be instantly generated upon sets of raw data. This new methodology holds the key to a new big data era: an age where global understanding reigns.

Keywords

Crisis management, Big Data, Decision Support, Agility.

[1] **GéNéPi** is funded by the French national association for research. It considers using the innate granularity of the management levels in crisis context to better support coordination. It is affiliated with the MISE approach.

[2] **MISE** incorporates several PhD works and scientific projects. It encompasses all the collaborative processes' generation approach, from the data retrieval to the agility of the generated and chosen process.

INTRODUCTION

Natural disasters, such as nuclear crises, floods, earthquakes or tsunamis, are now coexisting with man-made disasters as industrial accidents or terrorist attacks. The innate features of such disasters multiply the number of services needed in a hurry, wherever their localization. And, the further the responders, the more civilians are involved in the crisis response. This increases both (i) the level of heterogeneity between the stakeholders, causing trouble with communication and synchronization; (ii) the volume of data to digest for the decision-makers, available in various shapes, speed, credibility, objectivity or accuracy. Then, in order to improve the response phase, how to support the decision-makers in making the stakeholders collaborate, as in getting a clear picture of the on-going situation?

Solutions answering this question have already been implemented. For example, the MISE project proposes to automatically deduce collaborative processes solving the crisis, from the analysis of situation models. However, it does not currently help the decision-makers dealing with the huge amount of emerging data. Hence, this paper aims to go further and proposes to deal with a new challenge: how to extract, analyze and process data in order to automatically obtain the crisis situations models?

A CASE STUDY ILLUSTRATION: THE LOIRE RIVER FLOODING

The use-case, introduced and described here, has been validated by an official French flood forecasting service. It will serve to check and answer the problematic just introduced.

500 million people concerned worldwide

500 million people worldwide have been, at least once, affected by a flooding in their life, and each year, floods take a 25.000 toll. In Europe, between 1998 and 2006, (Nabet, 2013) 25 billion euros have been lost, 500.000 people left their home and 700 people died. In France, 74% of the cities are exposed to this risk and 80% of the damage due to natural disasters is actually caused by floods.

The third most feared natural disaster in France

A mighty river

The Loire River is the longest river in France: it flows along 1.012 km and its drainage basin covers 30% of the French territory. A mixed Loire flooding, caused by both oceanic and Cévenol precipitations, is at the third place of the most feared natural disasters in France for it has a long, wide and destructive progression.

A complex crisis response

One of the official flood simulation focuses on the Middle Loire, between Nevers (35,000 inhabitants) and Angers (149,000 inhabitants), impacted by a “1/50 per year” flood - with 2% chance of happening per year. Such a flooding sets in motion: 6 counties, 2 defense zones, one ministry, (CEPRI, 2014) 23 different kinds of ‘official’ stakeholders and nine crisis cells. These cells are set up to centralize decisions and facilitate the collaboration on the field.

Several needs emerging from the field

The Figure 1 shows the geographical and administrative boundaries existing in the Middle Loire area, along with the evolution of the expected flooding. This one lasts for seven days, of which five are on red alert. On the fourth day, all the vals[3] have to be partially evacuated at the same time. This represents almost 400 kilometres of river and surroundings to cover for the stakeholders, where several sensitive buildings and expose networks are at risks.

To manage the information sharing and the collaboration, on this geographical area, in that timeframe, assistance in defining, monitoring and modifying the collaborative behaviour is needed.

[3] Val: polder or incised valley protected from small floods by dikes

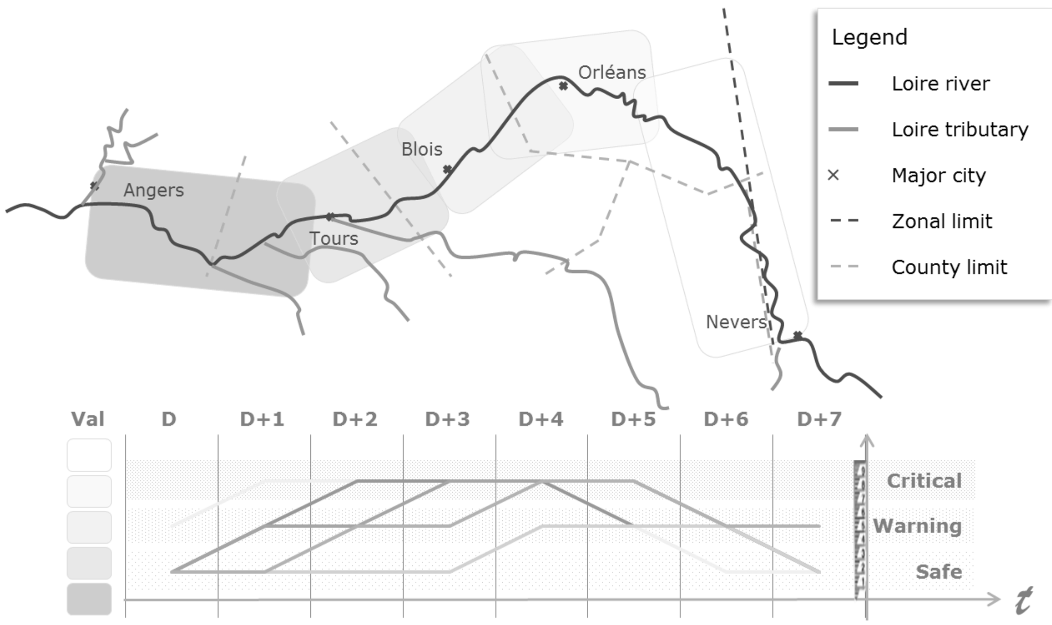

Figure 1 – Geographical and administrative view of the Middle Loire Area, along with the simulation of water levels during a 1/50 per year Loire flooding

Besides, the more time passes by, the more available the data. Consequently, the volume of information to infer and transfer for the stakeholders increases. Nevertheless, the overall digitalization of the space is the opportunity to give accuracy to the response process. For example, new data sources, (Van der Aalst, 2014) as newsfeeds, social media, sensors, mobile devices, etc., could provide information: (i) on the crisis' characteristics, as a new danger of landslide; (ii) on the available resources, as the future arrival of doctors on the field; (iii) on the environment, as the updated evolution of water level.

To help the stakeholders take advantage of big data (more data or information about the situation), without its disadvantage (too much data without filter or aggregation), the automation of the data management is needed.

Thanks to the preciseness of the use-case presented here, two key issues have been found. If answered, the decision-makers will have time saved and a better awareness of the situation to take the -critical- decisions left.

THE AUTOMATION OF THE COLLABORATIVE BEHAVIOUR

First, in order to manage the information sharing and the collaboration among the stakeholders, MISE proposes to instantly define, monitor and update the collaborative behaviour.

MISE (Bénaben, 2012) involves the definition, the architecture and the deployment of a Mediation Information System, which aims to automate the design, the orchestration and the supervision of collaborative processes. In the case of crisis management, this approach of decision support is effective: (i) during the preparation phase with the data pre-collection and (ii) during the response phase with the design of the situation model representing the partners, the context and the objectives.

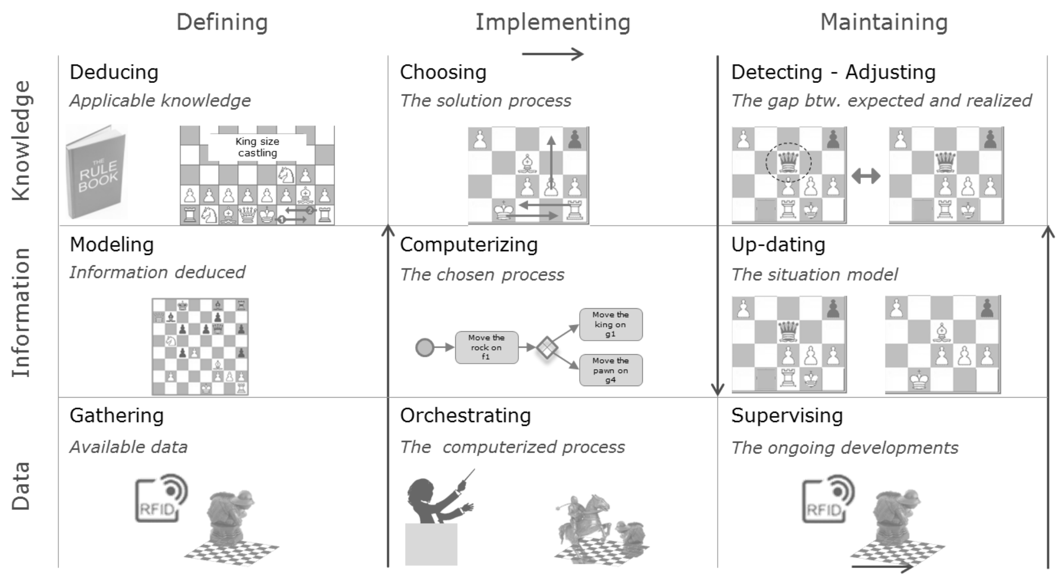

Figure 2 – The functioning of the Mediation Information System supporting the stakeholders' collaboration during a crisis situation (MISE project research works), illustrated by a chess game.

The Figure 2 illustrates schematically this principle and clarifies it thanks to a chess match analogy. The columns shows: the definition process for the situation model; its transformation into a response process and a workflow; the agility management: when the gap between the attended and real situation is too important, the MIS loops and defines the three situation models again. The rows correspond to (Ackoff, 1989): **the data level,** formed with observations, events and characteristics of the environment; **the information level,** needing relational connections to contextualize and give meaning to the data; **the knowledge level,** requiring the description of additional connections, as business rules, to enrich the information.

Thanks to this solution, the collaboration is reinforced by the definition, the orchestration and the monitoring of collaborative processes; the stakeholders, involved in a crisis situation, are now able to synchronize their actions without further efforts.

THE AUTOMATION OF BIG DATA MANAGEMENT: FROM RAW DATA TO KNOWLEDGE

Secondly, to help the stakeholders take advantage of big data, GéNéPi offers a methodology to automate the data management life-cycle, in order to improve the accuracy and pertinence of the situation model.

A new way to facilitate decision making

The big data management, also called **big data** (Kalyvas, 2015), is a process that facilitates the decision-making, through a swift analysis of large amounts of data, of different types, from a variety of sources, to produce a stream of usable knowledge (Power, 2014). Its usefulness depends on the distance between the context and the result of the big data analysis (Demchenko, Grosso, De Laat, 2013): if one person considers a data as information, another person, in another situation, could conversely ignore the same data.

Four famous big data features

The definition of big data above introduces the four main features, on which several researchers agree:

- **The volume** refers to the quantity of data generated continuously per time unit. It is captured, stored, to be recoverable and analyzed (Power, 2014; Demchenko, Grosso, De Laat, 2013; Hashem, Yaqoob, Anuar, 2015; Krishnan, 2013);
- **The variety** refers to the uncontrollable diversity of data types (videos, pictures, numbers, ...) and data format (structure, unstructured, ...) from both known and unknown sources (Power, 2014; Demchenko, Grosso, De

Laat, 2013; Hashem, Yaqoob, Anuar, 2015; Krishnan, 2013; Ohlhorst, 2012; Raghupathi, Raghupathi, 2014);

- **The velocity** refers to the speed with which the data is produced and the speed needed to process the data in time (Power, 2014; Demchenko, Grosso, De Laat, 2013; Hashem, Yaqoob, Anuar, 2015; Krishnan, 2013);
- **The veracity** refers to the trustfulness, the objectivity, the authenticity, and the security surrounding data. (Demchenko, Grosso, De Laat, 2013; Ohlhorst, 2012; Lukoianova, Rubin, 2014)

The particularity of crisis management concerns its lack of boundaries; conversely, industrial entities are able to control the volume, variety and veracity of their own data. GéNéPi aims to take advantages of this lack for the decision-makers benefit.

The available solutions managing one or several big data features

The tools currently developed worldwide (Chalmers, Bothorel, Clemente, 2013) include solutions for at least one of these features. For example, the volume can be tackled (Grolinger, Hayes, Higashino, 2014) by Map Reduce[4]. Similarly, the variety can be avoided thanks to the existence of metadata (Krishnan, 2013), as it identifies the data content. Though, there is no single solution that would take into account the veracity as Lukoianova and Rubin (Lukoianova, Rubin, 2014) define it: the objectivity, the truthfulness, and the credibility of the data need to be controlled, especially in a crisis context.

The methodology to obtain a model from big data analysis by managing all the big data features

The challenge is to deduce a model that will facilitate the decision-making inside a crisis cell, with the good level of preciseness and the good scope (Kaisler, Armour, Espinosa, 2013). As an example, in (Barthe-Delanoë, Truptil, Bénaben, Pingaud, 2014) a certain level of alpha or beta particles linked to a certain wind speed should generate an event instance such as: "the level of contamination exceeds the maximum threshold".

The Figure 3 presents the imagined principle, within the GéNéPi project, to automate this process. The solution treats all the features of big data in one run, thanks to five common steps. These steps involve the collection, homogenization and transformation of raw data; in addition, a certain level of agility is ensured by the loop step to optimize the performance of the solution.

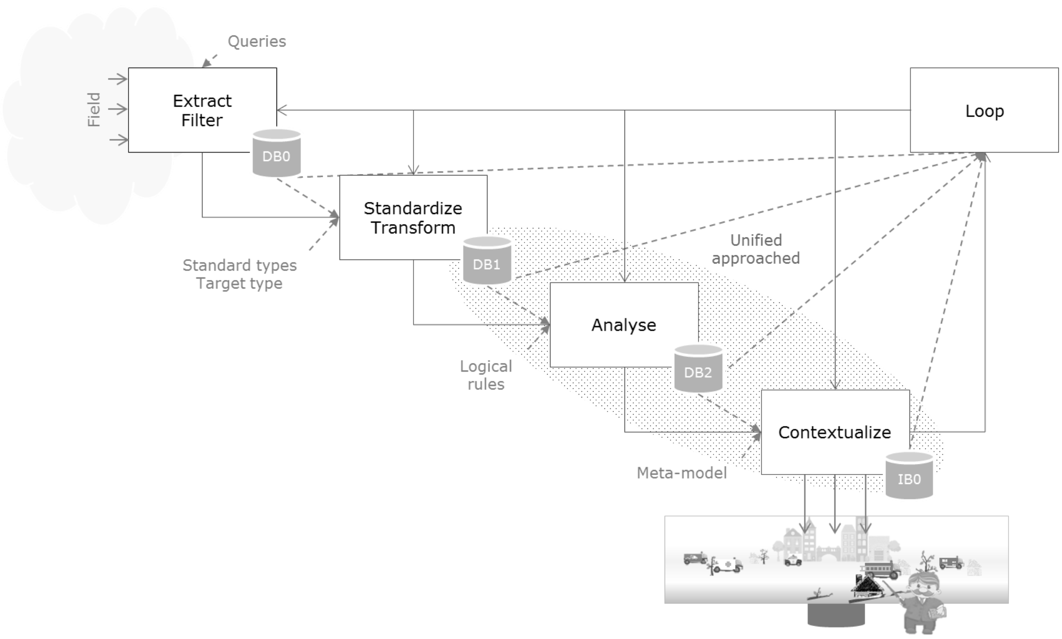

Figure 3 - The five steps supporting the data life cycle, from the data retrieval to the model representation of the knowledge

[4] **Map reduce**: method that permits to process extremely large amount of data by requisitioning numerous servers

Step 1: The Data Retrieval and Data Filtering

During a crisis management large amount of data are available from, for example: (i) the stakeholders, who are involved in the crisis response along with their sensors, their observations and their experiences, (ii) people, who want to improve the crisis response with their individual knowledge and skills, (iii) the victims, who want to communicate in order to be found, (iv) unknown sources through open data or the new internet of things (Van der Aalst, 2014). The choices of aims and sources are important at this stage to limit the volume, as well as (Kaisler, Armour, Espinosa, 2013) an addictive accumulation of data.

Once data is gathered from trusted sources, it needs to be filtered. This step enables (Xu, Huang, Jiang, 2014) the management of the data velocity and data volume. The goal is to reduce (Kaisler, Armour, Espinosa, 2013) the storage capacity of DB0 (cf. Figure 3). Here, the acceptance level of data into the system will depend on the money invested in the data storage: the higher the investment, the more accurate the generated models.

Step 2: The Data Type Standardization and Transformation

To avoid all the issues linked to the variety (Kaisler, Armour, Espinosa, 2013), the data has to be pre-treated in order to identify its format and type during the analytic phase. The unreadable data can be removed, while the remainder can be sorted into families (picture, text, photo, video, etc.), tagged and turned into a standardized family type.

Then, another operation transforms the data of these new types into a unique target type. All the data will then be interoperable thanks to a **unified approach** of interoperability (Chen, Doumeingts, Vernadat, 2008). This means that the target type, linking the data with one another, is not imposed on the owners (**integrated approach**), but is used to match one format with another.

Step 3: The Unified Data Set Analysis

To prepare the transformation of data into information, sets of linked data have to be judged, as any new information should be, in terms of objectivity, truthfulness and credibility. For this purpose, the system needs logical rules, in the same way as our brain needs connections. It has to be noted that the rules used in MISE have been deduced from field observations. A v-level (veracity level) can also be set thanks to (Lukoianova, Rubin, 2014) tools that detect subjectivity, opinions, biases, deception and credibility. Here, the possibility to re-use data is also a necessity (Demchenko, Grosso, De Laat, 2013): without memory, our brain cannot improve itself.

Step 4: The Contextualization

At last, in order to obtain information from sets of data and connections, a meta-model serves as a model in the modeling process. The MISE project has already defined one (Mace-Ramète, Lamothe, Lauras, Bénaben, 2012). The idea is to instantly deduce the instances of the situation model framework, so that the instances gathered in IB0 (cf. Figure 3) can be easily compared or easily retrieved.

In GéNéPi, the situation model as the situation models to-be, simulating decisions’ outcomes, are concerned by this methodology.

Loop: The Management of Several Evaluation Criteria

The solution described above must now include iterative methods (Demchenko, Grosso, De Laat, 2013) to continuously improve the quality of the data, information and knowledge stored into the system. A loop on each previous step, and from one to another, allows the system to control the volume, the velocity and the veracity by (Kaisler, Armour, Espinosa, 2013) adding, updating, searching for, retrieving, removing or tagging the items. This last step fosters the continuous analytics offered by the solution.

This improvement of the original solution shows that big data is nothing but a great opportunity for the decision-makers involved in a crisis response. Thanks to it, the models representing the crisis situation will gain accuracy and foresight.

CONCLUSION

The study of a particularly feared disaster in France enables the GéNéPi's researchers to identify two of the decision-makers' needs: (i) help in managing the information sharing and collaboration; (ii) help in managing the available data.

Knowing this, MISE offers to support decision-making, during a crisis, thanks to the definition, orchestration and monitoring of a collaborative process. In addition, the GéNéPi project works to automate the creation of the situation model. For the first time, a methodology proposes to control all the issues due to big data management in once. Thanks to it, the geographical scope of the decision makers will be able to spread, instantly taking into account new data sources to fit their needs. And, the more numerous the sources, the more the situation model could be accurate, for the benefits of both the decision-makers and the collaborative processes.

This approach has now to be developed and technological choices have to be made: how will the data be extracted, filtered, unified and contextualized, in regards with the required aspects described in this article?

REFERENCES

1. Nabet, F. (2013) - Etude du réajustement du lit actif en Loire moyenne, bilan géomorphologique et diagnostic du fonctionnement des chenaux secondaires en vue d'une gestion raisonnée - Doctoral dissertation, Paris 1.
2. CEPRI (2014) - L'évacuation massive des populations - Guide du CEPRI, Europe.
3. Van der Aalst, W. M. (2014) - Data scientist: The engineer of the future - *Enterprise Interoperability VI*, Springer International Publishing, pp. 13-26.
4. Bénaben F. (2012) - Conception de Système d'Information de Médiation pour la prise en charge de l'Interopérabilité dans les Collaborations d'Organisations – HDR, Institut National Polytechnique de Toulouse.
5. Ackoff R. L. (1989) - From data to wisdom - *Journal of applied systems analysis*, 16, 3-9.
6. Kalyvas, J. R. (2015) - A Big Data Primer for Executives – In Big Data, *CRC Press.*
7. Power D. J. (2014) - Using 'Big Data'for analytics and decision support - *Journal of Decision Systems*, vol. 23, no 2, p. 222-228.
8. Demchenko Y., Grosso P., De Laat C., et al (2013) - Addressing big data issues in scientific data infrastructure - *Collaboration Technologies and Systems (CTS),* IEEE World Congress, p. 48-55
9. Hashem I. A. T., Yaqoob I., Anuar N. B., et al (2015) - A Big Data Primer for Executives - In Big Data, *CRC Press.*
10. Krishnan K. (2013) - Data Warehousing in the age of big data - Elsevier *Science & Technology Books*.
11. Ohlhorst F. J. (2012) - Big data analytics: turning big data into big money - *John Wiley & Sons.*
12. Raghupathi W., Raghupathi V. (2014) - Big data analytics in healthcare: promise and potential - *Health Information Science and Systems*, vol. 2, no 1, p. 3.
13. Lukoianova T., Rubin V. L. (2014) - Veracity roadmap: Is big data objective, truthful and credible? - *Advances in Classification Research Online*, vol. 24, no 1, p. 4-15.
14. Chalmers, S., Bothorel, C., Clemente, R. P. (2013). Big Data-State of the Art.
15. Grolinger K., Hayes M., Higashino W. A. et al (2014) - Challenges for mapreduce in big data - *Services (SERVICES)*, IEEE World Congress, p. 182-189.
16. Kaisler S., Armour F., Espinosa J. A., et al. (2013) - Big data: Issues and challenges moving forward - *System Sciences (HICSS),* 46th Hawaii International Conference, p. 995-1004.
17. Xu L., Huang Z., Jiang H., et al. (2014) - Providing flexible file-level data filtering for big data analytics.
18. Barthe-Delanoë, A. M., Truptil, S., Bénaben, F., & Pingaud, H. (2014) - Event-driven agility of interoperability during the Run-time of collaborative processes - *Decision Support Systems*, 59, 171-179.
19. Chen D., Doumeingts G., Vernadat F (2008) - Architectures for enterprise integration and interoperability: Past, present and future - *Computers in industry*, 59(7), 647-659.
20. Mace-Ramète, G. M., Lamothe, J., Lauras, M., & Benaben, F. (2012, June) - A road crisis management metamodel for an information decision support system - *Digital Ecosystems Technologies (DEST), 6th IEEE International Conference,* (pp. 1-5)