HAL
open science

Génération d'un modèle de situation en temps réel : application à la gestion de crise

Audrey Fertier, Anne-Marie Barthe-Delanoë, Aurelie Montarnal, Sébastien Truptil, Frederick Benaben

To cite this version:

Audrey Fertier, Anne-Marie Barthe-Delanoë, Aurelie Montarnal, Sébastien Truptil, Frederick Benaben. Génération d'un modèle de situation en temps réel : application à la gestion de crise. MOSIM'18 - 12ème Conférence internationale de Modélisation, Optimisation et SIMulation, ISAE; IMT Mines Albi, Jun 2018, Toulouse, France. 8 p. hal-01852079

HAL Id: hal-01852079

https://imt-mines-albi.hal.science/hal-01852079

Submitted on 22 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GENERATION D'UN MODELE DE SITUATION EN TEMPS REEL : APPLICATION A LA GESTION DE CRISE

Audrey FERTIER[1], Anne-Marie BARTHE-DELANOË[2], Aurélie MONTARNAL[1], Sébastien TRUPTIL[1], Frédérick BÉNABEN[1]

[1]Centre Génie Industriel, Université de Toulouse, Albi, France
audrey.fertier@mines-albi.fr,
aurelie.montarnal@mines-albi.fr,
sebastien.truptil@mines-albi.fr,
frederick.benaben@mines-albi.fr

[2]Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France
annemarie.barthe@ensiacet.fr

RESUME : *Chaque jour, de plus en plus de sources de données sont rendues accessibles, et ce à l'échelle mondiale. Ceci constitue une opportunité sans pareil pour les organisations d'améliorer leur connaissance de la situation (« situational awareness ») dans un environnement de plus en plus instable et imprévisible. Cette connaissance sert de base à la prise de meilleures décisions et ce dans un laps de temps plus court qu'auparavant. Le projet GéNéPi a vu le jour afin d'aider cette prise de décision, dans un contexte multi-acteurs. Il vise à faciliter la collaboration inter-organisationnelle, en temps réel, quelque soit la complexité ou l'instabilité de l'environnement collaboratif. Afin de définir et tester la solution proposée par GéNéPi, un cas d'étude de gestion de crise a été développé en partenariat avec des institutionnels. Cet article présente une nouvelle approche pour automatiser la création et la mise à jour d'un modèle représentatif de la situation, en temps réel, afin de : (i) fournir aux décideurs une vue opérationnelle commune, (ii) établir un cadre de définition et de maintien des processus collaboratifs au cours du temps afin de veiller à la bonne poursuite des objectifs du réseau collaboratif.*

MOTS-CLES : *Modélisation, Big Data, Connaissance de la situation, Gestion de crise, Vue opérationnelle commune.*

1 INTRODUCTION

Durant une situation de crise, l'interconnexion des réseaux contraint fortement les différents acteurs en charge de la réponse à la crise (services de secours, institutions, opérateurs de réseau d'électricité, d'eau, de télécommunications, etc.). Afin de rendre la collaboration possible entre ces différents acteurs, des cellules de crise sont établies. Elles opèrent au niveau opérationnel, comme aux niveaux décisionnels. Le but principal de ces cellules inclut le partage d'information et la prise de décision (centralisée). Ceci nous amène aux constats suivants. D'une part, la gestion des multiples informations entrantes et sortantes de la cellule de crise monopolise une grande partie du temps des acteurs de la cellule. Par ailleurs, la nature instable et imprévisible de la situation complique la tâche des décideurs : bien que physiquement éloignés du terrain, ils doivent avoir la meilleure connaissance possible de la situation, en temps réel.

L'accroissement des sources de données disponibles durant la dernière décennie représente une opportunité sans pareil d'obtenir rapidement des informations sur la situation courante. Malheureusement, si les cellules de crise ne veulent pas être submergées par une avalanche de données brutes et d'informations, il leur faut limiter leurs recherches aux capteurs et appareils déjà identifiés et mis en place au préalable sur le terrain, ou aux rapports internes.

Dans ce contexte, afin de venir en aide aux différentes parties prenantes présentes sur le terrain de la collaboration et afin d'appuyer les décideurs dans leur travail, le projet GéNéPi ambitionne de concevoir, orchestrer et surveiller en temps réel la réponse à une crise (vue comme un ensemble de processus collaboratifs). Un des points innovants présenté par le projet est la génération automatique d'un modèle de la situation collaborative, alimenté via les données brutes émises par le terrain. Ce modèle est créé et mis à jour en deux étapes : (i) un modèle issu de la vérité terrain décrit les faits provenant directement de sources de données connues et fiables ; (ii) de nouvelles informations sont déduites sur la base de cette vérité terrain, en s'appuyant sur une base de connaissances et à l'aide de règles de déduction.
Durant ces deux phases, le système instancie un méta-modèle unique, en tant que « spécification explicite d'une abstraction » (Bézivin and Gerbé 2001). Dans le cas du projet GéNéPi, un méta-modèle est utilisé pour décrire la collaboration des nombreuses parties-prenantes pris dans la réponse à une crise. Ceci est un cas complexe de situation collaborative en milieu instable et imprévisible, avec une contrainte temporelle forte.

Ce papier présente une nouvelle méthodologie pour aider les décideurs dans la gestion des données et de l'information, au sein des cellules de crise françaises. Le résultat consiste en un modèle de situation, qui sert d'entrée à : (i) la déduction et la surveillance des processus collaboratifs ; (ii) la génération d'une vue opérationnelle commune, ou COP (Common Operational Picture), mise à disposition pour chaque cellule de crise.

La seconde section de l'article s'intéressera aux opportunités et aux menaces que représente l'avalanche de données disponibles lors d'une situation de crise. La troisième section montrera comment en tirer profit compte tenu des caractéristiques particulières à une situation de crise. La quatrième section s'intéressera aux événements qui seront à prendre en compte durant la réponse à la crise. Enfin, la dernière section présentera deux approches permettant de répondre à la question détaillée dans la troisième section, dont une nouvelle méthodologie, adaptée aux caractéristiques des évènements décrits dans la quatrième section.

2 UNE QUANTITE EXCESSIVE DE DONNEES DISPONIBLES A TRAITER EN TEMPS REEL

D'après Ackoff (Ackoff 1989), la donnée représente « les propriétés des objets et des événements », l'information est « contenue dans les descriptions » (Qui ? Quoi ? Où ? Combien ?) et la connaissance est « transmise par les instructions » (Comment ?). Cette section explique d'où proviennent les données disponibles, pourquoi il est nécessaire de les traiter dans leur ensemble, et à quelle vitesse elles doivent être traitées par les cellules de crise.

2.1 Les besoins des décideurs

Les réseaux, qu'ils soient sociaux, de capteurs ou d'organisations, vont croissants et génèrent de plus en plus de données. Ces sources de données sont nombreuses et hétérogènes. Elles introduisent une complexité dans la gestion de crise à plusieurs niveaux.
Si les décideurs veulent tirer parti de cette nouvelle situation, ils devront: localiser les sources de données; les identifier; déduire de nouvelles informations décrivant la situation de crise; résumer et partager leurs informations et leurs connaissances à l'intérieur de la cellule de crise; etc. Leur objectif est d'obtenir la vision la plus fiable et la plus précise possible de la situation de crise, en tant qu'image opérationnelle commune (i.e. COP, Common Operational Picture), afin de pouvoir prendre des décisions en toute connaissance de cause (Dautun and Lacroix 2013).

2.2 La contrainte temporelle

Une connaissance partagée de la situation (que l'on retrouve nommée « shared situational awareness » dans la littérature) peut être définie comme « the degree to which team members have the same SA on shared SA requirement » (Endsley and Jones 2001). Cette connaissance (et son partage) doit être mise à jour en temps réel pour s'adapter à l'instabilité intrinsèque de la situation de crise. Pour la générer, les données et les informations doivent être collectées auprès des parties prenantes impliquées dans la réponse à la crise. Malheureusement, il faut un temps important et précieux pour obtenir et diffuser les bonnes données (ou informations) après de la bonne personne –ou du bon appareil, du bon capteur– au bon moment. C'est la raison pour laquelle les décideurs choisissent volontairement de limiter le nombre de leurs sources de données.

2.3 L'équilibre entre rapidité et qualité

Ces limites volontairement posées permettent aux cellules de crise de gagner du temps dans leur processus de prise de décision, mais elles soulèvent néanmoins plusieurs problèmes:

- **Une perception datée de la situation**. En situation de crise, l'instabilité est la norme, ce qui rend les choses presque impossibles à prévoir. Il est donc essentiel d'être informé à tout moment d'une quelconque évolution de la situation. Les parties prenantes impliquées dans une crise s'attendent à ce que les systèmes respectent cette exigence de temps réel et jugent sévèrement les échecs. Par exemple, l'application pour smartphones SAIP (Système d'Alerte et d'Information des Populations) (Ministère de l'Intérieur 2016a) s'est avérée inutile durant l'attentat terroriste qui a frappé la ville de Nice au soir du 14 juillet 2016. En effet, l'application a relayé l'alerte avec un retard de trois heures et n'a été publiée qu'à 1h34 dans la nuit du 14 au 15 juillet (Ministère de l'Intérieur 2016b).
- **L'erreur de jugement**. Les plans d'urgence des cellules de crise françaises comprennent des réunions planifiées régulièrement. L'un des objectifs de ces réunions est le partage de l'information pour initier les prises de décisions régulières. Néanmoins, l'ordre d'arrivée des données peut impacter le processus de prise de décision. Par exemple, les observations de: (a) un mouvement de foule ; (b) quelques personnes blessées ; (c) des bruits soudains et forts ; peuvent être interprétés de plusieurs façons suivant la séquence de ces observations. Dans la séquence (a, b, c), la situation peut être comprise comme un mouvement de foule classique. Dans le cas de la séquence (c, b, a), la situation peut être interprétée comme un mouvement de foule conséquent à des coups de feu.
- **La perte d'information**. Certaines données, tels que les signaux faibles, n'attirent pas l'attention par nature. D'autres données quant à elles ne correspondront pas (en terme de format et ou de sémantique) au système d'information des cellules de crise. De fait, toutes ces données sont ignorées et définitivement perdues.

En résumé, le nombre croissant de données disponible représente une opportunité unique d'améliorer la connaissance de la situation en temps réel, au sein des cel-

lules de crise, tout en facilitant la prise de décision à travers une COP toujours plus précise. Dans le même temps, les acteurs pris dans la réponse à la crise doivent faire face aux problèmes liés aux données elles-mêmes, comme les délais, les erreurs de jugement ou les pertes. Pour les éviter, les systèmes d'informations dédiés à la gestion de crise sont nécessaires pour aider les décideurs en temps réel (Finn et al. 2015) et générer le meilleur modèle possible de la situation afin d'obtenir la COP attendue.

Comment générer automatiquement un modèle de situation décrivant la situation collaborative en cours et comment le mettre à jour en temps réel pour répondre à l'instabilité inhérente à une situation de crise ?

3 LA GENERATION D'UN MODELE DE SITUATION EN TEMPS REEL

L'exploration de données (data mining) peut être considérée comme une réponse pertinente au problème de la génération automatique d'un modèle de situation. Les approches d'exploration de données sont plutôt dédiées aux entreprises qu'aux situations de crise, et il existe de grandes différences entre ces deux champs d'application. Cette section vise à mettre en lumière ces différences et à définir comment l'exploration de données pourrait être utilisée pour gérer la grande quantité de données de crise disponibles, compte tenu des difficultés soulignées dans la section 2.

3.1 Les particularités du traitement de données dans l'industrie

Les techniques de data mining sont fréquemment utilisées par les entreprises pour surveiller les relations clients (CRM), ou pour identifier les meilleures opportunités de marché (BI), avec pour objectif final une aide à la prise de décision. H. Wang et S. Wang (Wang and Wang 2008) soulignent qu'une approche répandue consiste à utiliser le data mining pour relier (i) la collecte de données, l'accès aux données et l'analyse des données et (ii) la création et l'exploitation de la connaissance. Dans ce contexte, trois hypothèses sont faites :

- **Les sources de données sont connues au préalable**. Les experts métier, en connexion avec les data miners, planifient les extractions de données en fonction de leurs objectifs. Ceux-ci étant connus en avance, il est finalement plus aisé de choisir les sources les plus intéressantes et d'analyser les flux de données avec des règles précises.
- **Une portée réduite des décisions**. Dans la plupart des cas, les entreprises ont décomposé leurs fonctions de façon précise. Par conséquent, la portée des actions qui peuvent être prises pour résoudre des problèmes ou atteindre un objectif est restreinte à un périmètre étroit.
- **Une grande fenêtre temporelle**. L'extraction des données peut être effectuée sur une période de temps importante afin d'améliorer l'exactitude et la pertinence des résultats.

3.2 Les particularités du traitement de données pendant la réponse à la crise

Les techniques de data mining dépendent de leur champ d'application. Par exemple, elles sont dirigées par les objectifs dans l'industrie, alors que dans le cadre d'une gestion de crise, elles sont dirigées selon les sources de données disponibles : le nombre et la diversité des sources étudiées tendent à diminuer les manques, les pertes d'information, les malentendus ou les jugements sans discernement. Cela représente un énorme volume qui doit être traité en temps réel, quelle que soit la variété des données, qui peuvent arriver sous forme d'images satellites, de photographies, de rapports, de tweets, d'articles, d'emplacements GPS ou de données capteurs. Si le traitement de ces données était automatisé, alors il serait possible d'inférer de précieuses connaissances pour compléter la COP représentant la situation de crise, c'est-à-dire une carte localisant les personnes, les acteurs de la réponse, les risques, les faits (Norheim-Hagtun and Meier 2010), les sentiments (Caragea et al. 2014) ou les activités en cours de réalisation.
Les situations de crise sont assez éloignées des cas industriels et de leurs hypothèses de travail :

- **Des sources de données floues et inconnues**. Les data miners ne peuvent pas prédire les résultats de leur travail, ni l'emplacement des données selon leur niveau de pertinence ou de précision. Ils ne peuvent pas non plus accéder à des données qui ne sont pas connues au préalable, tout en ayant pour objectif d'en étudier le plus grand nombre (afin de fiabiliser le résultat).
- **Des utilisations imprévisibles**. Une donnée peut faire référence à plusieurs types d'informations, liées ou non aux objectifs du décideur. Par conséquent, les règles de classification doivent être à la fois permissives, pour traiter le plus de données possibles, et précises, pour répondre aux attentes des décideurs.
- **Une évaluation de la performance difficile à prévoir**. Le nombre d'activités disponibles pour répondre à un risque ou pour traiter un effet néfaste augmente avec le nombre d'acteurs dans la cellule de crise (tous niveaux opérationnels et stratégiques confondus). Par ailleurs, il n'est pas possible d'établir des indicateurs de performance en avance de phase, et les résultats du processus de réponse à la crise sont difficilement évaluables sur le court terme.
- **Une fenêtre temporelle étroite**. Un seul événement est suffisant pour déclencher une crise en très peu de temps ou en modifier les caractéristiques majeures. Dans ces conditions, le système d'information doit être capable de détecter de tels événements en quelques minutes.

3.3 Un système informatique à trois niveaux servant de base à la gestion des données en situation de crise

Les particularités de l'industrie et de la gestion de crise révèlent deux différences importantes portant sur: (i) le volume et la qualité des données qui peuvent être extraites en temps réel; (ii) le nombre des actions possibles à entreprendre. C'est pourquoi il est intéressant d'utiliser ces grands flux de données, non seulement pour maintenir un modèle de la situation, mais aussi pour définir et surveiller les actions à entreprendre. Dans cette perspective, un système d'information à trois niveaux est nécessaire:

- **Le niveau « données ».** Les données brutes sont directement extraites de sources de données hétérogènes et variées (telles que Systèmes d'Information Géographiques, capteurs, réseaux sociaux ...). Les données extraites sont hétérogènes, nombreuses, plus ou moins fiables et datées.
- **Le niveau « information »**. Ce niveau correspond à la contextualisation des données précédentes. Concrètement, le niveau d'information vise à fournir un état de la situation de crise en temps réel. La COP ainsi obtenue consiste en une carte où les instances importantes et pertinentes pour la cellule de crise sont localisées. Une fois mis en œuvre, ce niveau pourrait soutenir la prise de décision à l'intérieur de chaque cellule de crise. Un méta-modèle est utilisé pour standardiser et faciliter cette modélisation.
- **Le niveau « connaissance »**. Il s'agit des activités à mener compte tenu de la situation de crise caractérisée par le niveau « information ». Les actions font par exemple partie du processus collaboratif qui est (i) déduit par les partenaires de GéNéPi et (ii) exécuté par les parties prenantes. Le processus de sélection des activités offre un large éventail de possibilités.

Les sections suivantes se focalisent sur les solutions pour traiter en temps réel les données émises par le théâtre de la crise et obtenir les informations les plus pertinentes afin d'alimenter le modèle de situation.

4 PARTICULARITES DES EVENEMENTS EMIS PAR UNE CRISE

L'événement peut être compris comme "la chose qui est arrivée" (comme défini par (Chandy and Schulte 2009; Etzion and Niblett 2010; Luckham and Schulte 2011). Par exemple, il peut s'agir d'un tsunami, d'un niveau d'eau atteignant un seuil, de la réception d'un courriel, de la nouvelle version d'une page Web, d'un commentaire sur un statut Facebook, etc. Lorsqu'un événement se produit sur le terrain ou est sur le point de se produire, son abstraction est créée dans un système d'information (Luckham and Schulte 2011) sous la forme d'un objet, d'un message ou d'un tuple. Dans cet article, l'événement sera considéré comme étant l'événement « informatisé » et il contiendra au moins:

- Un identifiant unique,
- L'heure de l'événement terrain (i.e. dans le monde physique),
- Une description de l'événement terrain (données ou informations sur le terrain).

Pour générer automatiquement le modèle de situation représentant la collaboration, le système doit décrire le contexte, les partenaires et la crise elle-même. Le contexte et les partenaires peuvent tous deux être représentés pendant la phase de préparation. C'est pourquoi le système pourrait, dans un premier temps, se concentrer sur la modélisation des risques et des effets de la crise.

4.1 Description générale des événements terrain à observer

Pour pouvoir éviter les erreurs ou les retards dans la génération de la COP à partir du modèle de situation, les événements terrain doivent être observés via l'Internet des Evénements (IoE). Il consiste en quatre classes telles que définies par Van Der Aalst (Van der Aalst 2016) :

L'Internet des Objets. Cette classe intègre tout objet physique connecté au réseau via une connexion Internet ou grâce à des étiquettes d'identification uniques telles que des codes à barres ou des étiquettes RFID. Pendant une crise, les systèmes d'identification, comme le «système d'information Internet sans fil pour la réponse médicale aux catastrophes», sont utilisés pour signaler la position des victimes et le dossier médical pendant le triage (Koenig and Schultz 2010).

L'Internet des Personnes. Ces données sont générées à partir des interactions sociales entre les personnes. Il peut s'agir d'e-mails, de messages envoyés via les réseaux sociaux tels que Facebook, Twitter ou Instagram. Pendant une crise, les gens ont tendance à se concentrer sur les médias sociaux pour communiquer ou pour être informés sur la crise en cours, en raison du principe « 1-to-n »: poster un seul message sur Twitter et n abonnés seront avertis en même temps. Ce crowdsourcing émergent est particulièrement intéressant dans le contexte de la gestion de crise pour soutenir la prise de conscience du contexte (Meier 2015).

L'Internet des Lieux. Cela inclut toutes les données avec une dimension spatiale. Par exemple, les tweets, les publications sur Facebook et même les images peuvent intégrer des attributs de géolocalisation fournis par les smartphones et les appareils photo modernes. Lors d'une crise, un pourcentage significatif (49% selon Kouadio et Douvinet (Kouadio and Douvinet 2015)) des victimes potentielles dispose d'un smartphone et peut être localisé sur une carte. Cette représentation de crise peut donc être utilisée comme une COP pour soutenir la prise de décision locale au sein des cellules de crise.

L'Internet du Contenu. Il représente les informations créées par les humains, comme les pages Web, Wikipedia, YouTube, les flux d'actualités. Pendant une crise, les recommandations officielles pour aider les victimes

pourraient être transmises directement aux personnes sur place qui veulent aider, etc.
Comme l'a souligné van der Aalst (Van der Aalst 2014), les données sur les événements constituent la principale source d'information. Tous ces événements disponibles sont nombreux et les données et informations qu'ils contiennent sont plus ou moins fiables, proviennent de sources variées, de différents types et formats, et sont datées.

4.2 Problèmes dus au volume, à la variété, à la vélocité et à la véracité des données terrain

Les événements disponibles décrivent des données et des informations provenant du champ de la crise, des cellules de crise de niveau supérieur et inférieur ou d'Internet. Cette section décrit les caractéristiques de ces entrées hétérogènes, également appelées *4V du Big Data* :

Le volume. Il se réfère à la quantité de données générées en continu par unité de temps. Le volume est proportionnel à la taille des données (Demchenko et al. 2013) et donc au type de données (Krishnan 2013). Un équilibre doit être trouvé entre l'élargissement de la zone d'influence et l'énergie dépensée à traiter les données à temps. Pourtant, plus il y aura de données, plus la découverte d'informations et de comportements cachés sera importante pour les décideurs (Hashem et al. 2015). À mesure que le temps passe, le volume de données disponibles continue d'augmenter tandis que la capacité disponible pour les traiter reste constante (Fan and Bifet 2013).

La variété. Elle fait référence à la diversité des types de données, comme les vidéos ou les journaux de données (Chalmers et al. 2013; Fan and Bifet 2013; Hashem et al. 2015; Krishnan 2013; Ohlhorst 2012; Power 2014), et les formats de données, en tant qu'ensembles structurés ou non structurés (Demchenko et al. 2013; Hashem et al. 2015). Les problèmes de contournement liés à la variété sont possibles grâce aux métadonnées identifiant ce qui est contenu dans les données réelles (Krishnan 2013), même si beaucoup de détails sont perdus dans ce processus.

La vélocité fait référence à la vitesse nécessaire pour récupérer et traiter les données dans le temps(Krishnan 2013; Power 2014). Pour éviter de perdre des données, qui pourraient être essentielles pour inférer des informations critiques aux décideurs, il est crucial d'améliorer la rapidité de toutes les étapes de l'analyse (Kaisler et al. 2013). La vélocité fait également référence à la vitesse à laquelle les données sont générées (Demchenko et al. 2013; Power 2014).

La véracité se réfère à la fiabilité, l'exactitude, la cohérence et la sécurité des données (Demchenko et al. 2013; Gantz and Reinsel 2011). Lukoianova et Rubin (Lukoianova and Rubin 2014) identifient également l'objectivité, la véracité et la crédibilité comme trois dimensions complémentaires de la véracité. (Leskovec et al. 2014) vont plus loin en affirmant que, compte tenu de l'évolution permanente des situations étudiées tout domaine considéré, le processus d'agrégation de données hétérogènes doit être réalisé en temps réel afin de pouvoir fournir une vue pertinente de la situation.

5 APPROCHES DE GENERATION DE MODELE DE SITUATION A PARTIR DE DONNEES BRUTES

D'après les caractéristiques énoncées ci-dessus, un bon outil d'analyse doit accéder à autant d'événements que possible (volume, vitesse), à partir de sources hétérogènes (variété). Il doit également calculer la véracité et la valeur de toutes les données à l'intérieur du système. Des méthodologies couvrant au moins un de ces points existent déjà. Cette section en présente deux.

5.1 Les approches générales traitant a minima un des 4V

KDD (Knowledge discovery in databases) par (Fayyad et al. 1996). Le processus de découverte de connaissances dans des bases de données structurées a été mis en place pour cartographier des données de bas niveau, trop volumineuses pour être facilement agrégées. Le processus utilise des modèles pour lier des données ensemble. Ces modèles doivent être validés (véracité) et potentiellement utiles (valeur). Ce processus tient en cinq étapes : sélection ; pré-traitement ; transformation ; data mining ; schémas d'interprétation et évaluation.

SA (Situational Awareness) par (Endsley 1995). La conscience de la situation d'un système est mesurée par sa capacité à percevoir des éléments dans un environnement particulier, à les comprendre et à les projeter dans le futur. L'ensemble du processus est composé de trois étapes principales (Fayyad et al. 1996):

- Perception des éléments: observer et récupérer le statut, les attributs et la dynamique des éléments pertinents sur le terrain.
- Compréhension de la situation actuelle: recueillir des éléments récupérés et déduire de nouveaux éléments qui véhiculent de plus en plus de significations pour les utilisateurs et leurs objectifs.
- Projection du statut futur: prévoir l'évolution du statut des éléments pour projeter sa future dynamique dans l'environnement.

L'objectif est de fournir les connaissances requises par les décideurs sous contrainte de temps (vélocité), sans perdre de vue de leurs objectifs (valeur).

Ces deux méthodologies couvrent partiellement les caractéristiques des données présentées précédemment (4 V du Big Data).

5.2 Une nouvelle approche prenant en compte les 4 V du Big Data

Le projet GéNéPi intègre nativement la gestion des données massives au système d'aide à la collaboration. Les événements sont directement analysés pour générer des modèles à jour, durant l'exécution des processus collaboratifs. Dans le cas d'une situation de crise, ils peuvent faciliter la déduction et l'orchestration de la réponse à la crise mais aussi fournir une COP à chaque cellule de crise.

Du point de vue de l'intelligence artificielle, le système GéNéPi peut être vu comme un agent (Russell and Norvig 2009): il peut percevoir son environnement, à

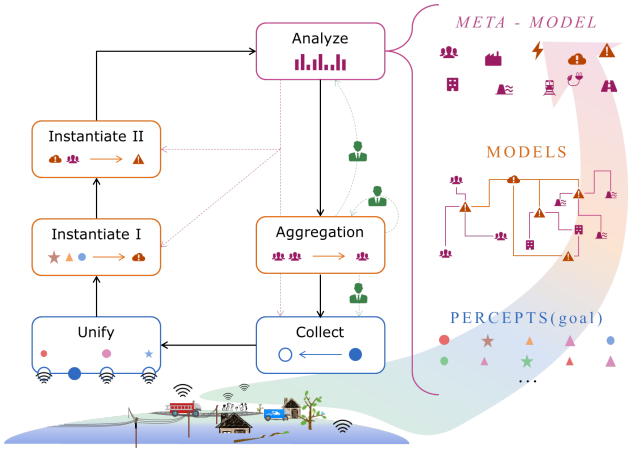

Figure 1. Les modules composant le système d'information à trois niveaux de GéNéPi.

travers la réception des percepts, et agir sur l'environnement, à travers l'orchestration d'actions coordonnées. Dans le cadre du projet GéNéPi, l'Interpreter GéNéPi est la brique logicielle responsable de la prise de conscience de la situation.

La structuration des informations. L'Interpreter GéNéPi illustre la situation de crise en cours en instanciant un méta-modèle (commun à tous les modèles de situation). Le méta-modèle (Bénaben et al. 2016) est composé de deux types de concepts:

- La première partie du méta-modèle peut être instanciée principalement avant la crise, lors de la phase de préparation. Cela concerne les capacités des parties prenantes, désireuses d'être impliquées ou prévues dans une procédure ; et les enjeux dans le domaine. Plus tard, lors d'une réponse à une crise, des partenaires inconnus peuvent émerger et de nouveaux enjeux peuvent être identifiés et le modèle peut être complété en temps réel ;
- La deuxième partie du méta-modèle concerne la crise elle-même : le danger, les risques et les événements affectant le territoire.

L'objectif de l'Interpreter GéNéPi. Le module Collect (voir Figure 1) détermine régulièrement l'objectif de l'Interpreter. Il indique quels aspects de l'environnement doivent être considérés. Les percepts - décrivant l'environnement de crise et perçus par le système - sont ainsi toujours adaptés aux priorités actuelles. Par exemple, au début de la réponse, les cellules de crise les plus basses dans la hiérarchie essaient d'abord de détecter les conséquences et les risques liés à la crise. Grâce à elles, les cellules de crise de niveau hiérarchique supérieur pourront évaluer l'ampleur de la crise. Dans ce cas, le module Collect doit prioriser la collecte de données pouvant être utilisées pour instancier la seconde partie du méta-modèle. Cette étape de *perception* est critique (Endsley 1995) compte tenu de la quantité de données à traiter dans le temps.

L'interopérabilité de tous les composants de l'Interpreter. Les données et informations sont reçues par le module Unify. Les informations reçues de sources fiables et connues peuvent être automatiquement ajoutées au modèle de situation tandis que les autres sont redirigées vers le module Instanciation (voir Figure 1). Au sein de l'Interpreter, des problèmes se posent rapidement en raison de la variété des données. C'est l'un des plus grands obstacles à la compréhension d'un grand volume de données (Kaisler et al. 2013). Pour contrer cela, le module Unify transforme automatiquement les données d'entrée en documents XML. Ces nouvelles données sont des événements informatiques. Ils comprennent (* obligatoirement): * une date de création ; * un identifiant ; * un sujet ; une géolocalisation ; et enfin une description de ou des événements observés sur le terrain et rapportée par la donnée ou l'information reçue par le module Unify.

Des événements bruts aux modèles de situation. Le module Instantiate vise à comprendre la signification des événements informatiques: il relie des données disparates pour développer (Endsley 1995) les connaissances que le système a de la situation courrante (« *Situation awareness* »). Ce module est divisé en deux parties. Instancier I, reçoit les événements provenant du module Unify et les transforme en instances du méta-modèle. Instantiate II déduit de nouvelles instances directement du modèle de situation. Par exemple, lorsqu'Instantiate I déduit un nouveau danger d'inondation à partir de l'évolution prédite de la rivière, Instantiate II ajoute un nouveau risque pour chaque bâtiment sensible menacé dans la zone de danger.

$$\forall\, Enjeu,$$
$$Risque = Danger * Vulnérabilité(Danger)$$

Un danger menace tous les enjeux qui lui sont vulnérables dans son périmètre. Par exemple, un risque de victime apparaîtra pour chaque foyer situé en zone inondable et un risque de coupure électrique apparaîtra pour chaque transformateur en zone inondable.

L'étape de compréhension. L'un des objectifs du module Analyze est de prévoir le futur, sous la forme d'un nouveau modèle, dérivé du modèle de situation, appelé

modèle projeté. Il est produit en combinant l'avancée du processus de réponse en cours, les progressions naturelles (prévisibles) de la crise et le modèle de situation courant. Par exemple, il peut prendre en compte l'évolution du niveau d'eau et du débit d'eau, prévue par le Service de Prévision des Crues. L'objectif est d'anticiper les divergences entre la situation en cours, représentée par la COP, et la situation qui devrait résulter de la réponse en cours, pour pouvoir adapter au plus tôt la réponse des autorités. Cette étape de *compréhension* permet au système d'atteindre le dernier niveau de conscience (« *Situation awareness* ») décrit par (Endsley 1995).
L'autre objectif du module Analyze est de surveiller la qualité (Wang and Strong 1996) des percepts, les modèles et le méta-modèle manipulés par l'Interpreter GéNéPi. Pour réussir, le module Analyze cherche à confirmer les résultats, à combler les informations manquantes ou à améliorer la véracité des modèles. Ces itérations permettent également d'alerter l'administrateur lorsqu'une source de données ou une règle d'exploitation doit être remise en question.

La visualisation des informations. La COP illustre à la fois la situation et les modèles projetés. Le but (Endsley 2015) est de communiquer efficacement sur «ce qui est entrain de se passer» et sur «ce qui risque de se produire ensuite », tout en réduisant de manière significative les incertitudes et le stress liés aux prises de décision (Dautun and Lacroix 2013).

Le point de vue utilisateur. Les modèles contiennent tellement d'informations que l'utilisateur doit pouvoir accéder à une granularité et un horizon géographique précis. Par exemple, au lieu du nom, de la fonction et de la géolocalisation de chaque victime potentielle, la cellule de crise de la zone de défense préférerait connaître le nombre total de victimes sur sa zone. C'est pourquoi, le module Aggregate a été configuré pour ajuster, en temps réel, la COP à chaque cellule de crise. Typiquement, pour la crue de la Loire, cela représente neuf points de vue différents (Fertier et al. 2016). Pour réussir, la COP est affichée dans une forme familière aux parties prenantes: une interface de SIG (Système d'Information Géographique) (voir Figure 2).
Le module Aggregate gère également les droits d'accès et les demandes d'informations complémentaires. Par exemple, si une information a été classée au niveau national, une autorisation formelle sera requise de tout maire, préfet ou préfet de zone qui voudrait y accéder.

6 CONCLUSION

Les décideurs impliqués dans la réponse à une crise n'ont pas le temps de prendre en compte de nouvelles sources de données, alors que ce serait l'occasion d'avoir une vue plus complète de la situation courante. Pour les aider, le projet GéNéPi propose à la fois de soutenir la collaboration et la gestion des données dans les cellules de crise. La solution pourrait être tirée du monde de l'industrie, si le monde de la gestion de crise n'avait pas ses propres hypothèses de travail: sources de données inconnues, utilisation imprévisibles des données, nombre d'actions possibles infini, temps passé à la collecte et l'analyse de données compté. Dans ce contexte, deux approches ont été étudiées et évaluées en considérant le nouveau volume de données à gérer au sein des cellules de crise. La combinaison de ces deux approches a permis la mise en œuvre d'un nouvel outil, ancré : (i) dans la suite d'outil existante IO Suite (Bénaben et al. 2016) qui aide déjà à la collaboration de toutes les parties-prenantes à la réponse à la crise ; et (ii) dans le prototype GéNéPi concurrent à la solution IO Suite. Les différentes règles (métier, transformation de modèle) ont été rédigées avec l'aide du CEREMA, en sa qualité d'expert métier.

A l'avenir, un tel outil, permettant la génération automatique d'un modèle de situation en temps réel, quelle que soit la complexité et l'instabilité du contexte de décision, bénéficiera non seulement aux décideurs pris dans une réponse à la crise, mais aussi :

- Aux industriels désireux de s'adapter rapidement aux variabilités de l'approvisionnement et/ou de la demande, tout en accédant aux informations cachées dans les données qui leurs sont maintenant accessibles en grand nombre (Agostinho et al. 2016) ;
- A toutes les personnes en prise avec des tableaux de bords de plus en plus complexes qui représentent directement la couche « données », et non la couche « information » comme peut le faire l'Interpreter GéNéPi.

Pour la suite, les utilisations de la solution IO Suite ou du prototype GéNéPi devraient permettre la constitution d'une base de connaissances regroupant les modèles de plusieurs situations de crise, à différent moment de la réponse. Si elle est utilisée, cette base de connaissance permettra à l'Interpreter GéNéPi d'apprendre à déduire des relations de causes à effets via un raisonnement à partir de cas (« *Case based reasoning* ») (Bennani et al. 2017) ou des règles seuils au bénéfice du module Instantiate I.

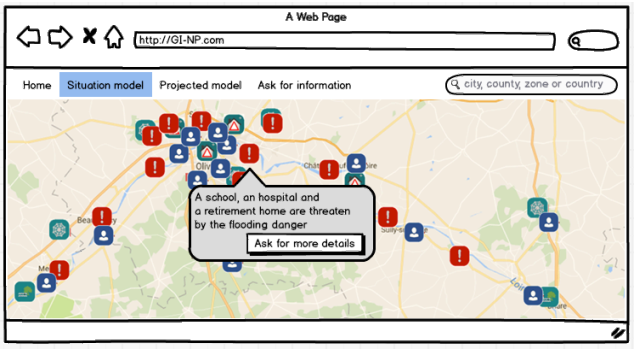

Figure 2. Une des interfaces utilisateurs, présentant le modèle de la situation.

REMERCIEMENTS

Les auteurs tiennent à remercier les membres du projet GéNéPi pour leur aide et leurs conseils. Les travaux

présentés dans ce papier sont financés par l'Agence Nationale de la Recherche via le projet GéNéPi (référence : ANR-14-CE28-0029).

REFERENCES

Ackoff, R. L. (1989). "From data to wisdom." *Journal of applied systems analysis*, 16(1), 3–9.

Agostinho, C., Pereira, J., Lorré, J.-P., Ghimire, S., and Benazzouz, Y. (2016). "A Distributed Middleware Solution for Continuous Data Collection in Manufacturing Environments."

Bénaben, F., Lauras, M., Truptil, S., and Salatgé, N. (2016). "A Metamodel for Knowledge Management in Crisis Management." *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 126–135.

Bennani, S., Maalel, A., Ghézala, H. B., and Abed, M. (2017). "Towards a Decision Support Model for the Resolution of Episodic Problems Based on Ontology and Case Bases Reasoning: Application to Terrorism Attacks." *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 1502–1509.

Bézivin, J., and Gerbé, O. (2001). "Towards a Precise Definition of the OMG/MDA Framework." *16th IEEE International Conference Automated Software Engineering (ASE'2001)*, San Diego, USA, United States, 273–280.

Caragea, C., Squicciarini, A. C., Stehle, S., Neppalli, K., and Tapia, A. H. (2014). "Mapping moods: Geo-mapped sentiment analysis during hurricane sandy." *Proceedings of the ISCRAM 2014 Conference*.

Chalmers, S., Bothorel, C., and Picot Clemente, R. (2013). *Big Data - State of the Art*.

Chandy, K., and Schulte, W. (2009). *Event processing: designing IT systems for agile companies*. McGraw-Hill, Inc.

Dautun, C., and Lacroix, B. (2013). "Placer l'humain au cœur des crises." *Lettre d'information sur les risques et crises INHESJ*, (38), p10–19.

Demchenko, Y., Membrey, P., Grosso, P., and de Laat, C. (2013). *Addressing big data issues in Scientific Data Infrastructure*. Piscataway, NJIEEE9781467364034.

Endsley, M. R. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors*, 37(1), 32–64.

Endsley, M. R., and Jones, W. M. (2001). "A model of inter and intra team situation awareness: Implications for design, training and measurement. New trends in cooperative activities: Understanding system dynamics in complex environments." *CA*, 46–67.

Etzion, O., and Niblett, P. (2010). *Event Processing in Action*. Manning Publications Co., Greenwich, CT, USA.

Fan, W., and Bifet, A. (2013). "Mining Big Data: Current Status, and Forecast to the Future." *SIGKDD Explor. Newsl.*, 14(2), 1–5.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From data mining to knowledge discovery in databases." *AI magazine*, 17(3), 37.

Fertier, A., Montarnal, A., Barthe-Delanoë, A.-M., Truptil, S., and Bénaben, F. (2016). "Adoption of big data in crisis management toward a better support in decision-making." *Proceedings of Conference on Information System for Crisis Response And Management (ISCRAM 16)*.

Finn, R. L., Watson, H., and Wadhwa, K. (2015). "Exploring big'crisis' data in action: potential positive and negative externalities." *Proceedings of the ISCRAM 2015 Conference*.

Gantz, J., and Reinsel, D. (2011). "Extracting value from chaos." *IDC iview*, 1142(2011), 1–12.

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.

Kaisler, S., Armour, F., Espinosa, J. A., and Money, W. (2013). "Big Data: Issues and Challenges Moving Forward." *2013 46th Hawaii International Conference on System Sciences*, 995–1004.

Koenig, K. L., and Schultz, C. H. (2010). *Koenig and Schultz's disaster medicine: comprehensive principles and practices*. Cambridge University Press.

Kouadio, J., and Douvinet, J. (2015). "Les smartphones peuvent-ils aider à une meilleure remontée des données en cas de crue rapide pour améliorer les systèmes d'alerte?" *BSGLg*, 64, 57–68.

Krishnan, K. (2013). *Data warehousing in the age of big data*. Newnes.

Leskovec, J., Rajaraman, A., and Ullman, J. D. (2014). *Mining of massive datasets*. Palo Alto.

Luckham, D., and Schulte, W. R. (2011). "Event Processing Glossary - Version 2.0." EPTS.

Lukoianova, T., and Rubin, V. L. (2014). "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?" *Advances in Classification Research Online*, 24(1), 4.

Meier, P. (2015). *Digital humanitarians: how big data is changing the face of humanitarian response*. Crc Press.

Ministère de l'Intérieur. (2016a). "Lancement de l'application mobile SAIP." *http://www.interieur.gouv.fr/Archives/Archives-des-actualites/2016-Actualites/Lancement-de-l-application-mobile-SAIP*, <http://www.interieur.gouv.fr/Archives/Archives-des-actualites/2016-Actualites/Lancement-de-l-application-mobile-SAIP>.

Ministère de l'Intérieur. (2016b). "Mise en oeuvre de l'application SAIP le soir du 14 juillet 2016." *http://www.interieur.gouv.fr/Actualites/Communiques/Mise-en-oeuvre-de-l-application-SAIP-le-soir-du-14-juillet-2016*, <http://www.interieur.gouv.fr/Actualites/Communiques/Mise-en-oeuvre-de-l-application-SAIP-le-soir-du-14-juillet-2016> (Jul. 21, 2016).

Norheim-Hagtun, I., and Meier, P. (2010). "Crowdsourcing for crisis mapping in Haiti." *Innovations: Technology, Governance, Globalization*, 5(4), 81–89.

Ohlhorst, F. J. (2012). *Big data analytics: turning big data into big money*. John Wiley & Sons.

Power, D. J. (2014). "Using 'Big Data' for analytics and decision support." *Journal of Decision Systems*, 23(2), 222–228.

Russell, S., and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Pearson, Upper Saddle River.

Van der Aalst, W. M. (2014). "Data scientist: The engineer of the future." *Enterprise interoperability VI*, Springer, 13–26.

Van der Aalst, W. M. P. (2016). *Process Mining - Data Science in Action*. Springer.

Wang, H., and Wang, S. (2008). "A knowledge management approach to data mining process for business intelligence." *Industrial Management & Data Systems*, 108(5), 622–634.

Wang, R. Y., and Strong, D. M. (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, 12(4), 5–33.