

LES ÉTAPES DE MODÉLISATION D'UNE SITUATION DE CRISE DANS LA LITTÉRATURE

Our « Age of Anxiety » is, in great part, the result of trying to do today's jobs with yesterday's tools—with yesterday's concepts. [With yesterday's ideals.]

– Marshall McLuhan, *The Medium is the Massage : An Inventory of Effects*

Introduction

Suite à un événement majeur (Lagadec, 1994), les décideurs doivent rapidement recueillir des informations pertinentes, obtenues si possible selon un processus de recherche et d'analyse éprouvé, pour comprendre la « dynamique générale » de la crise.

Dans ce contexte, nous proposons de mettre au point un système d'information capable de générer un ensemble structuré d'informations décrivant la situation de crise en cours. Cet ensemble, présenté sous forme de modèle (Bézivin, 2005), doit alimenter directement le niveau 2 de conscience des décideurs, c'est-à-dire leur compréhension de la situation de crise, via une Common Operational Picture (COP) adaptée à chaque cellule de crise.

Le chapitre 1 nous a permis d'identifier notre problématique : *Comment collecter, puis interpréter des données brutes issues de sources hétérogènes pour déduire un ensemble d'informations structuré et utilisable, décrivant une situation courante, tout en gérant le volume, la variété, la vélocité et la véracité des données et des informations manipulées ?*. Le chapitre 2 cherche maintenant à identifier les méthodes qui composeront le système d'information recherché, pour répondre à la problématique posée. L'architecture doit permettre :

- la *collecte* des données « brutes » contenues dans l'Internet des événements ;
- l'*interprétation* des données collectées, c'est à dire leur mise en relation, pour déduire des informations permettant de se représenter la situation dans laquelle se trouve l'utilisateur ;
- la *structuration* des informations disponibles sur la crise en cours, pour obtenir un modèle utile et partageable de la situation de crise.

2. Les étapes de modélisation d'une situation de crise dans la littérature

tout en répondant aux défis liés aux 4Vs du Big data, définis à la section 1.3.4) :

- *volume* : comment sélectionner les données ou les informations de façon à optimiser leur taux d'utilisation en cellule de crise ?
- *variété* : comment harmoniser les formats et les types de données et d'informations manipulés ?
- *vélocité* : comment minimiser le temps nécessaire aux processus de collection, d'interprétation et de structuration ?
- *véracité* : comment vérifier les informations transmises aux cellule de crise ?

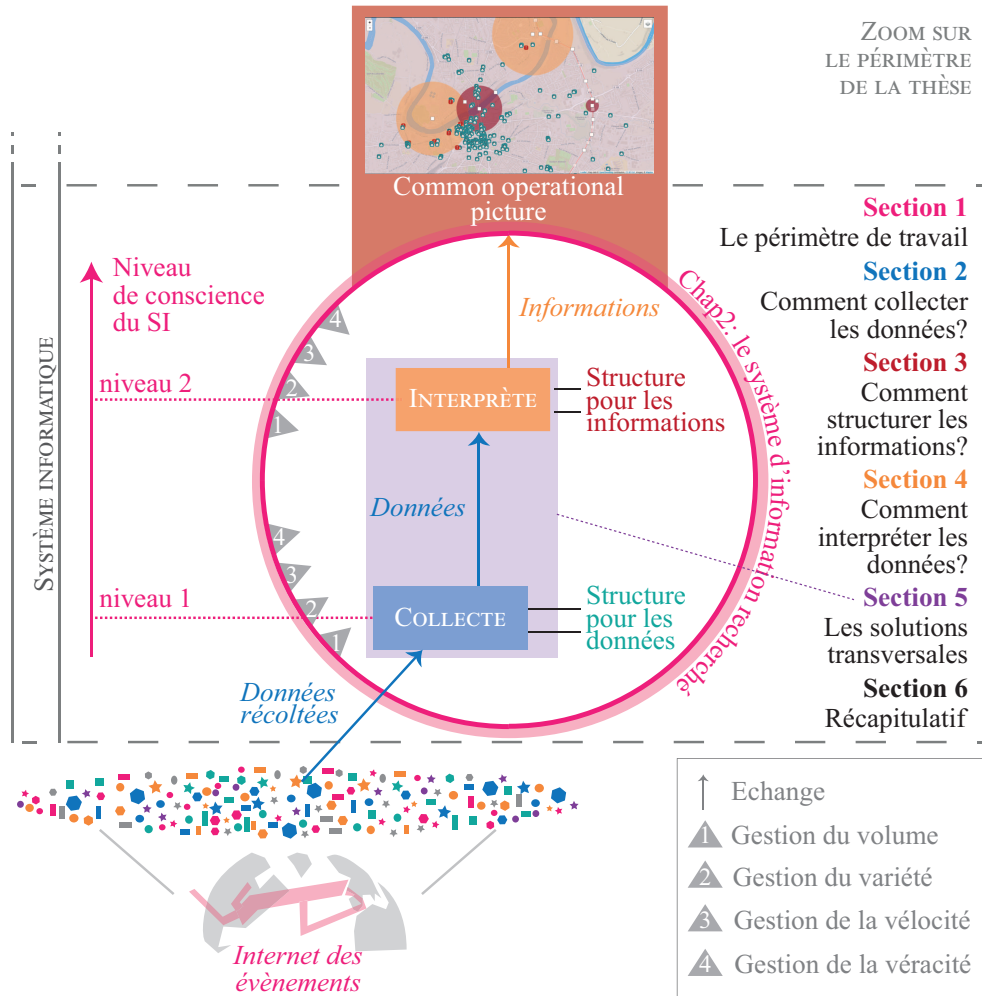


FIGURE 2.1 – Périmètre d'étude du Chapitre 2

La Figure 2.1 présente l'organisation du chapitre. La Section 1 commence par poser les hypothèses et le cadre de recherche suivi par les travaux décrits dans ce manuscrit. Les Sections 2, 3 et 4 cherchent respectivement à identifier les méthodes de collecte, de structuration et d'interprétation existantes dans la littérature. A chaque fois, les résultats seront évalués selon les contraintes liées à leur utilisation en situation de crise et dans un contexte de « Big data ». A la section 5, une recherche transverse aux états de l'art, présentés aux sections 2, 3 et 4, est proposée, pour tenter d'identifier des solutions hybrides présentes dans la littérature, comme par exemple, une méthode capable d'assurer à la fois la collecte et l'interprétation

de données brutes. Enfin, la Section 6 offre un récapitulatif des travaux les plus pertinents, compatibles avec notre domaine d'application et capables de répondre à notre problématique, compte tenu de nos hypothèses de départ.

2.1 Le cadre de recherche

2.1.1 Les hypothèses de travail

Un système d'information, destiné à collecter des données provenant de l'Internet des événements, pourrait détecter et identifier automatiquement des sources de données disponibles sur une zone géographique donnée. En gestion de crise, cette étape peut être réalisée en préfecture, durant la phase de préparation. Nous considérons donc pouvoir avoir accès à des sources fiables, existantes, identifiées par des experts métier. Nous posons ainsi notre première hypothèse :

Hypothèse 1. *Le système d'information sait, a priori, se connecter à des sources de données fiables, disponibles quelque soit le territoire étudié.*

Pour gérer le volume de données, plusieurs méthodes existent déjà, comme par exemple la technique connue sous le nom de Map-Reduce ([Zhang et al., 2015](#); [Dean et Ghemawat, 2008](#)) qui permet de séparer des données en plusieurs ensembles qui seront traités en parallèle. Ce constat nous amène à poser l'hypothèse ci-dessous qui signifie, en particulier, qu'aucune étude de scalabilité ou de limite de calcul n'est conduite dans ces travaux.

Hypothèse 2. *Le système d'information, supposé à ressources infinies, saura s'adapter à n'importe quel volume de données, grâce à une multiplication des moyens de calcul et à la parallélisation des étapes de collecte et d'interprétation.*

Pour gérer la variété des données, l'approche unifiée est envisageable, grâce à la standardisation du langage de description des données rendue possible par des communautés telles que le World Wide Web Consortium (W3C). Ce qui nous amène à une nouvelle hypothèse, décrite ci-dessous. Concrètement, le système d'information recherché ne collectera pas les données provenant directement de sources de données de « premier ordre », comme un capteur de mesure hydraulique, mais à des sources de données de seconde main, qui fournissent des données pré-traitées, telles que les relevés hydrométriques fournis par les Services de Prévision des Crues (SPC) français en cas de crue.

Hypothèse 3. *Le système d'information reçoit des données ou des informations déjà uniformisées via l'utilisation d'un langage de description unique.*

Enfin, nous supposons, comme indiqué par [Wolbers et Boersma \(2013\)](#) ou [Sophronides et al. \(2017\)](#), que les COPs sont efficaces pour communiquer rapidement et correctement des informations. Plus encore, nous considérons, comme relevé par [Stanton et al. \(2010\)](#), que la COP permet aux décideurs d'accéder à une conscience de la situation soutenue par un ensemble d'information compréhensible. Ainsi, comme proposé par l'hypothèse ci-dessous, l'amélioration de la conscience des décideurs est conditionné par l'amélioration des informations accessibles depuis la COP.

Hypothèse 4. *L'amélioration de la conscience de la situation du système d'information contribue directement à la conscience de la situation de son utilisateur.*

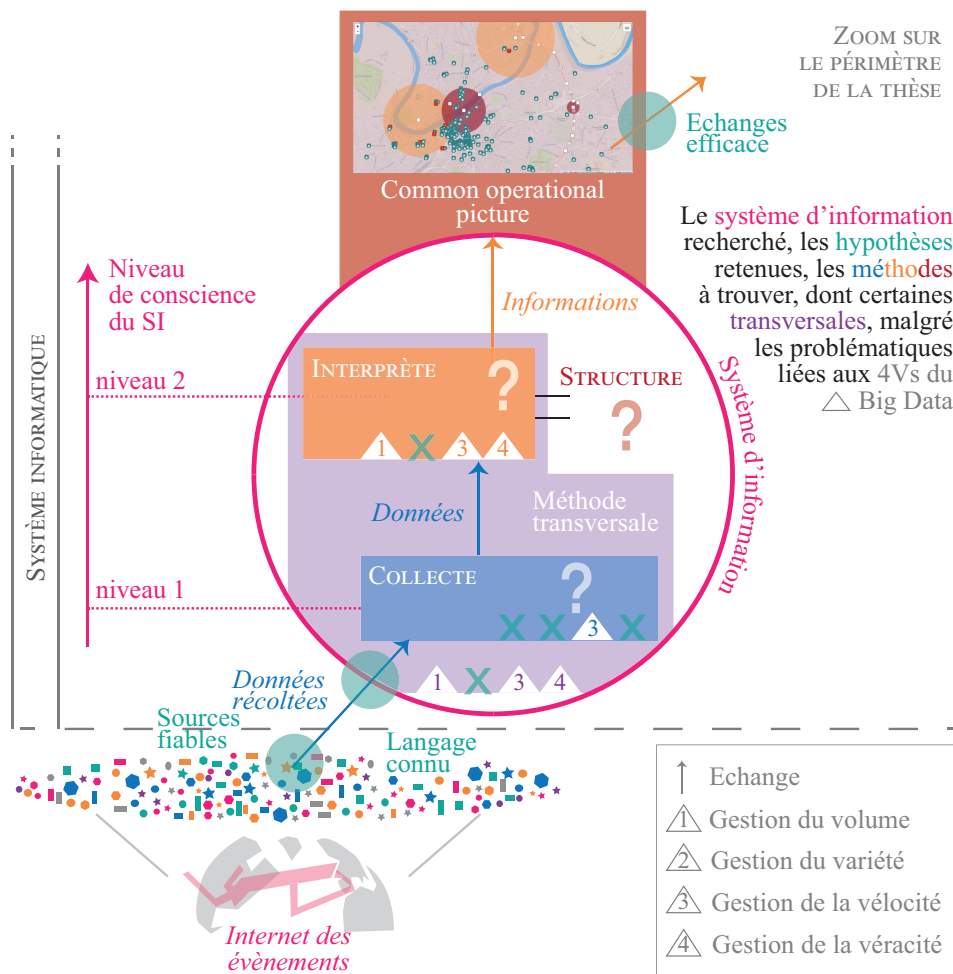


FIGURE 2.2 – Les hypothèses régissant notre cadre de recherche, présentées sur le cadre de réponse proposé pour mettre au point le système d'information recherché, mis en relation avec les numéros de sections du chapitre

La Figure 2.2 met en avant notre périmètre de recherche, encadré par nos quatre hypothèses :

1. les sources de données utilisées ont été jugées fiables, en phase de préparation ;

2. l'augmentation du volume de données en entrée n'impactera pas les performances du système d'information ;
3. les données disponibles au sein de l'Internet des évènements sont échangées selon un format commun et les informations manipulées par le système d'information sont structurées selon une structure de référence ;
4. les informations partagées avec les décideurs en cellule de crise sont comprises rapidement et entièrement.

2.1.2 La méthodologie suivie

La Figure 2.2 présente notre nouveau cadre de réponse, adapté à notre périmètre de recherche. A chaque étape, les conditions d'acceptation sont héritées des 4Vs du Big data encore à traiter.

Pour compléter ce cadre de recherche, par rapport à l'existant, nous suivrons la méthode décrite par Thomé et al. (2016) pour obtenir un état de l'art exploitable, qui consiste à :

1. formuler le problème et poser les frontières de la recherche souhaitée : la problématique a été posée à la fin du chapitre 1 et la frontière est délimitée par nos quatre hypothèses de recherche ;
2. poser les mots clefs et identifier les journaux concernés, sélectionner des articles selon leur résumé, puis selon leur contenu.
3. présenter les résultats de façon ordonnée et les mettre à jour régulièrement : les résultats seront représentés selon les trois étapes identifiées en orange et en bleu sur la Figure 2.2.

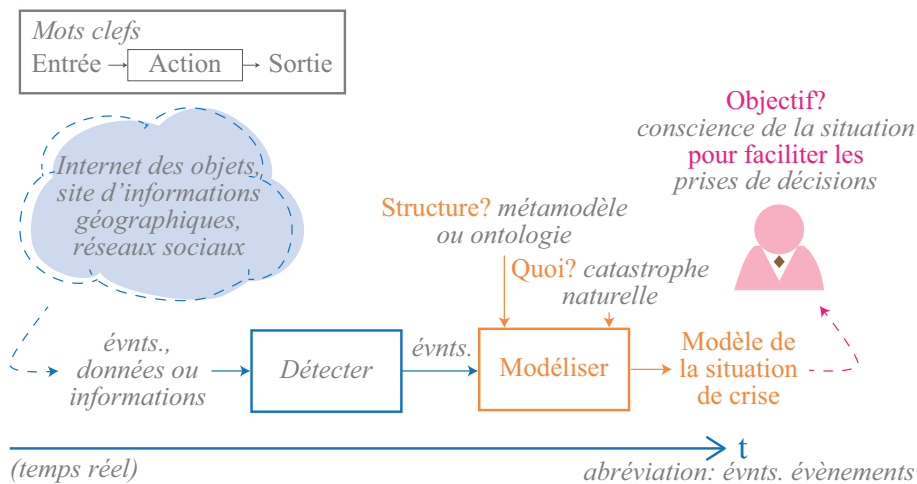


FIGURE 2.3 – Mots clefs choisis pour chacun des états de l'art réalisés dans ce chapitre, aux sections 2.2, 2.3 2.4 et 2.5

La Figure 2.3 présente les mots clefs sélectionnés pour chaque état de l'art. L'objectif était d'obtenir 15 à 50 articles par requête, pour ensuite n'en sélectionner qu'une dizaine. Il a donc fallu définir des mots clefs assez près de nos besoins pour obtenir 20 à 67% d'articles pertinents dans l'ensemble des résultats obtenus par requête. Trois requêtes ont ainsi été définies :

- la première requête recherche des méthodes de collecte capable de détecter, en *temps réel*, des évènements, des données ou des informations

provenant de diverses sources de données présentes dans l'Internet des objets, l'Internet des lieux (les *informations géographiques regroupées par des volontaires*), ou l'Internet des personnes (les *réseaux sociaux*) dans le contexte d'une situation de crise ;

- la seconde requête recherche des structures, telles que les ontologies ou les métamodèles, dédiées à la représentation de situation crise dues à des *catastrophes naturelles* ;
- la dernière requête recherche des méthodes d'interprétation capable de traiter des (événements) pour *modéliser* une situation courante, améliorer la *conscience de la situation* de décideurs et faciliter ainsi les *prises de décisions* ;

Enfin, les trois dernières requêtes, utilisées à la section 5, combinent les mots clefs ci-dessus, en élargissant nos critères de sélection, pour tenter d'identifier des méthodes transversales qui auraient pu nous échapper.

2.2 Les méthodes de collecte en temps réel

Pour commencer à compléter notre cadre de réponse, cette partie s'intéresse à l'étape de collecte des données, en bleu sur la Figure 2.2. Pour cette étape, l'objectif est d'éviter toute implication des décideurs lors de la collecte des données sur une crise en cours. Pour ce faire, le module *Collecte* a accès à des sources de données jugées fiables en phase de préparation (hypothèse 1) et qui mettent à disposition des données écrites sous un format unique, connu (hypothèse 3).

2.2.1 Les types de collecte

Selon Perera et al. (2014), les méthodes qui permettent, à distance, de récupérer des données appartiennent à deux familles distinctes :

- les méthodes qui interrogent, sur demande, une ou plusieurs bases de données via l'utilisation de requêtes écrites dans des langages dédiés.
- les méthodes qui souscrivent une fois à un abonnement particulier et reçoivent automatiquement les événements émis sous cet abonnement, quel que soit leur envoyeur.

La première méthode pourrait amener le système d'information à passer à côté de données critiques, rendues disponibles entre deux requêtes (Perera et al., 2014). A contrario, la seconde méthode est capable, selon Fabret et al. (2001), de gérer des centaines d'évènements par seconde. Celle-ci serait donc plus adaptée à la phase de réponse, où la dynamique de la crise contraint le temps disponibles pour les prises de décisions, et où chaque information peut jouer un rôle critique.

Les *Évènements* sont des objets informatiques (Luckham et Frasca, 1998), des données, *représentant* un événement réel, passé, en cours, prédit ou simulé. Ils doivent être constitués d'un identifiant, d'une date (d'émission, de réception ou d'observation) et d'une description. Ils peuvent aussi être géo-localisés. Les événements informatiques représentent des événements tels que la mesure de température issue d'un capteur, une alerte émise suite au dépassement d'un seuil

de mesure de radioactivité, ou encore la mise à disposition d'un rapport sur un serveur.

2.2.2 Les méthodes de collecte dans la littérature

Suite aux abonnements de notre système d'information à diverses sources de données, des flux d'évènements peuvent être reçus de façon quasi-continue. Pour cette raison, l'objectif est de *détecter* des évènements de valeur, et non pas de collecter tous les évènements disponibles. Pour rappel, le système d'information a accès à des évènements émis au sein de l'Internet des personnes, de l'Internet des objets, de l'Internet des contenus et de l'Internet des lieux (cf. section 1.3.4).

Pour classer les principales méthodes de détection utilisées aujourd'hui, nous avons interrogé la base documentaire *Web of Science* avec la première des requêtes décrites sur la Figure 2.3 : « (detect (event* OR data OR information) AND ("social media" OR microblog* OR "internet of things" OR "volunteered geographic information") AND real-time) ». Seuls les articles catégorisés sous la dénomination « Computer Science Information System » ont été retenus. En tout, 38 publications ont été obtenues, puis étudiées. Les résultats les plus pertinents sont résumés dans le Tableau 2.1.

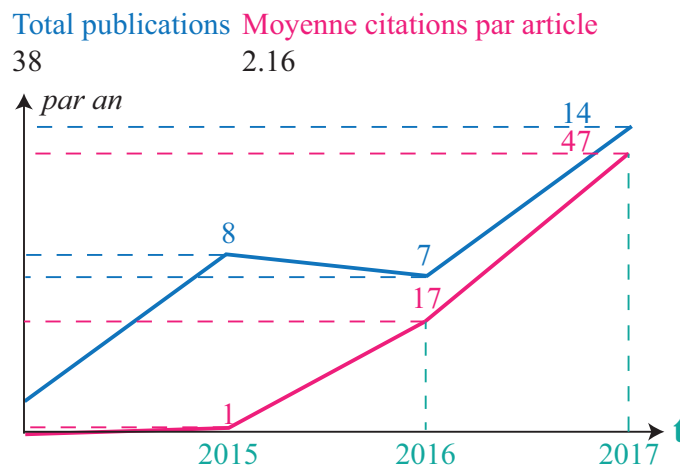


FIGURE 2.4 – Résultats obtenus sur *Web of Science*, des *Articles* publiés après 2014 et sous la catégorie *Computer Science Information System*, à la requête : (detect (event* OR data OR information) AND ("social media" OR microblog* OR "internet of things" OR "volunteered geographic information") AND real-time)

La figure 2.4 présente l'évolution du nombre de publications répondant à notre requête sur la base *Web of Science*. La détection de nouveaux évènements semble gagner en popularité chaque année depuis 2014. Pour rappel, [Luckham et Frasca \(1998\)](#) souhaitait déjà mettre au point un système capable d'extraire de l'information à partir d'évènements informatiques.

2.2.3 La collecte de données en situation de crise

Le Tableau 2.1 présente les résultats de l'état de l'art présenté à la Figure 2.4. Au vu des résultats détaillés dans le tableau, nous proposons une classification

2. Les étapes de modélisation d'une situation de crise dans la littérature

Réf.	Défi posé	Solution proposée	Couverture ?
Hu et Ni (2018)	(b) Détecter un type d'objet précis sur une bande vidéo	Filtrer les images par fréquence d'énergie	1 type de sources, 1 type d'objets
Akbar et al. (2017, 2018)	(b) Ingérer et analyser en temps réel des événements et des données historiques	Un moteur d'interprétation basé sur des règles seuils, associé à des techniques d'apprentissage	X types de sources, (?) types d'objets
Rathore et al. (2017)	(a) \cap (c) Détecter des événements à partir des flux de données générés sur les réseaux sociaux	Le classement des messages selon leur lieu, leur date d'émission et leur contenu	1 type de sources, X types d'objets
Wang et al. (2017)	(b) \cap (c) Détecter des événements de trafic grâce à des données GPS et un réseau social	La détection de motifs via l'utilisation de tenseurs représentant le trafic	X types de sources, 1 type d'objets
Xu et al. (2016, 2017)	(a) \cap (c) Récupérer des messages sur un réseau social pour une crise donnée	Un métamodèle et une analyse sémantique du contenu par rapport au lieu d'émission	1 type de sources, 1 type d'objets
Laylavi et al. (2017); Hua et al. (2016)	(a) \cap (c) Récupérer des données sur un réseau social, pour une crise ou un domaine donné	Le calcul de proximité entre un message sélectionné et un ensemble de termes spécifiques	1 type de sources, 1 types d'objets
Mallah et al. (2015)Car-rino et al. (2016)	(b) \cap (c) Détecter des situations dangereuses durant un évènement précis grâce à des volontaires	Le classement des personnes par groupe et l'estimation de la pression exercée	1 type de sources, 1 type d'objets
Musaev et al. (2015)	(a,b) \cap (c) Détecter un type d'incident grâce aux données provenant des réseaux sociaux ou de capteurs	Classification des messages sélectionnés et calcul de la probabilité de l'incident par zone	X types de sources, 1 type d'objets
Sugitani et al. (2015)	(a) \cap (c) Détecter des événements terrain sur un réseaux social	Le calcul de proximité entre les mots clefs pour chaque message émis du même endroit, pour une fenêtre de temps donnée	1 type de sources, X types d'objets
Chen et al. (2014)	(b) Détecter une attaque	La classification des événements entrants suivie de la détection d'anomalie	1 type de données, 1 type d'objets

TABLE 2.1 – Résultats les plus pertinents de l'état de l'art sur les méthodes dédiées à la détection d'évènements en temps réel, en vue de décrire une situation courante, à partir des données provenant de (a) l'Internet des personnes, (b) l'Internet des objets ou (c) l'Internet des lieux

des méthodes de collecte en quatre catégories, selon les types de données qu'elles acceptent en entrée et les types d'évènements qu'elle cherchent à détecter :

- les méthodes $1 \text{ à } 1$, où un type particulier de données connues est analysé pour détecter des types d'anomalies (Hu et Ni, 2018; Chen et al., 2014), ou isoler des groupes de données qui présentent des caractéristiques particulières (Xu et al., 2017; Laylavi et al., 2017; Carrino et al., 2016);
- les méthodes $x \text{ à } 1$, où des données provenant de plusieurs sources sont analysées pour détecter un évènement en particulier, comme des bouchons (Wang et al., 2017), ou des glissements de terrain (Musaev et al., 2015);
- les méthodes $1 \text{ à } x$, où un type de données connues est analysé pour détecter plusieurs type d'évènements comme l'analyse de tweets permettant de détecter des tremblements de terre ou des incendies à l'échelle d'un pays (Rathore et al., 2017), ou n'importe quel type d'évènement au niveau local (Sugitani et al., 2015).
- les méthodes $x \text{ à } x$, où des données, provenant de plusieurs sources, sont analysées pour détecter plusieurs types d'évènements.

En situation de crise, le système d'information recherché devra détecter en temps réel les conséquences de la crise en cours, les enjeux menacés ou touchés, ou encore les moyens à dispositions des cellules de crise. Il faut donc au minimum une méthode de collecte de type (1 à x). Mais, pour pouvoir suivre la dynamique (vélocité des données et informations) et les différents types de conséquences dues à la crise (variété des informations), une méthode de type (x à x) est indispensable.

2.3 Les méthodes pour structurer les informations disponibles

Pour continuer à compléter notre cadre de réponse, cette partie s'intéresse à l'étape de structuration des informations, en orange sur la Figure 2.2. Pour cette étape, l'objectif est de définir des concepts et de généraliser des relations liant tout ou partie des informations utilisables en cellule de crise. Comme proposé par Bourey et al. (2006), cette étape permettra surtout de résoudre les problèmes d'interopérabilité entre les systèmes d'information des parties-prenantes à la crise et entre les services composant notre système informatique.

2.3.1 Les types de structures

Deux types de structures prédominent lorsqu'elles sont utilisées par des systèmes d'information :

- le *métamodèle* présentant des concepts et un vocabulaire nécessaire à la modélisation cohérente d'un ensemble de situations pour un point de vue donné (Bézivin, 2005). Il peut lui-même être décrit selon un ensemble de concepts pré-définis alors appelé *métamétamodèle* (Karagiannis et Kühn, 2002);
- les *ontologies* construites dans le but de faciliter la réutilisation et le partage d'information (Uschold et Gruninger, 1996; Gómez-Pérez,

2. Les étapes de modélisation d'une situation de crise dans la littérature

2001), grâce à un ensemble de concepts, pour une partie du monde et un point de vue donnés. Dans cette catégorie, on retrouve souvent les taxonomies construites pour nommer des concepts similaires et les organiser entre eux, comme, par exemple, la taxonomie exhaustive des primates vivants (Groves, 2001).

Ces définitions héritent toutes les deux des recommandations écrites par (Gruber, 1995) sous le terme d'ontologie, pour encadrer un processus de conceptualisation. Elles sont similaires par leur conception, mais poursuivent des objectifs différents : (i) l'instanciation ou (ii) le rapprochement de concepts pré-définis.

De son côté, Karagiannis et Kühn (2002) se réfère à trois niveaux de modélisation précis : le métamodèle de (niveau 3) qui définit les concepts utilisés par ses métamodèles de (niveau 2) qui définissent eux-même les concepts destinés à être instanciés pour modéliser une situation particulière (niveau 1). Bézivin (2005) propose même d'utiliser un standard mis au point en 1997 par l'Object Management Group (OMG) pour unifier le niveau 3 de méta-modélisation : le Meta-Object Facility (MOF).

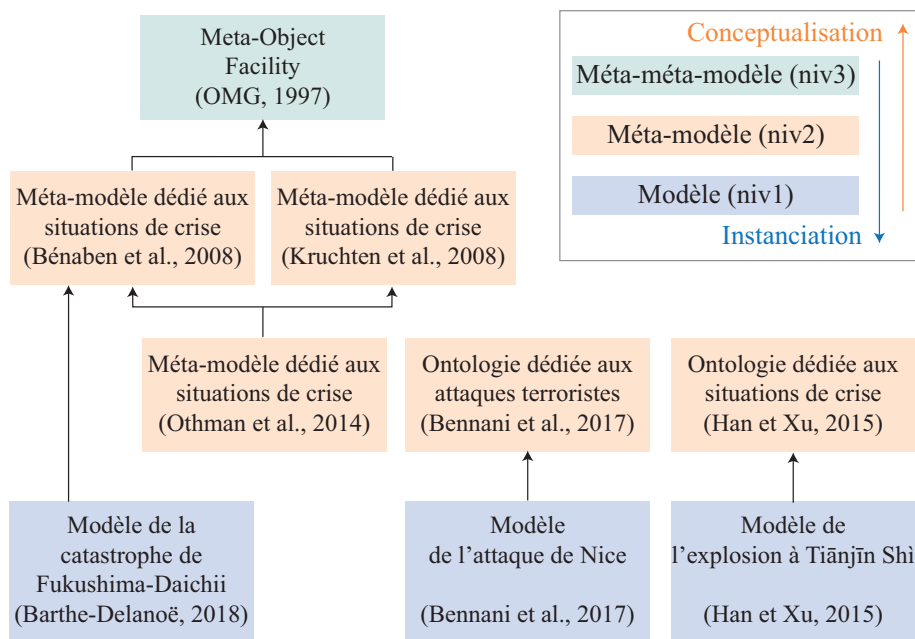


FIGURE 2.5 – Illustration du principe de méta-modélisation, inspiré de Karagiannis et Kühn (2002); Djurić et al. (2005) et illustré par les travaux de Othman et al. (2014); Costa et al. (2012); Han et Xu (2015); Barthe-Delanoë et al. (2018) et Bennani et al. (2017)

La Figure 2.5 illustre le fonctionnement et l'utilisation de ces trois niveaux de modélisation à partir d'exemples tirés des travaux en interne de notre laboratoire utilisés par exemple pour modéliser la crise nucléaire de Fukushima-Daiichi (Barthe-Delanoë et al., 2018), ou de travaux concurrents, comme ceux de :

- Othman et al. (2014) s'inspirent du métamodèle utilisé dans notre laboratoire et ajoutent quelques concepts tirés des travaux de Kruchten et al. (2008);

- [Bennani et al. \(2017\)](#) proposent une ontologie construite à partir de cas passés et dédiée à la réponse à des attaques terroristes ;
- [Han et Xu \(2015\)](#) proposent une ontologie formée de types d'évènements et de types de tâches adaptées pour y répondre. L'ontologie est utilisée pour suivre la crise qui suivit l'explosion d'un entrepôt de produit chimique à Tiānjīn Shì (Chine).

Tous les modèles de niveau 2 ne se rapportent pas systématiquement à un modèle de niveau 3.

2.3.2 Les méthodes pour structurer des informations dans la littérature

Pour structurer les informations décrivant tout ou partie de la situation de crise, seuls les métamodèles ou les ontologies de niveau 2, représentés en orange sur la Figure 2.5, nous intéressent. Pour identifier ceux dédiés à décrire des catastrophes naturelles, nous avons interrogé la base documentaires [Web of Science](#) avec la deuxième des requêtes décrites sur la Figure 2.3 : « ((ontology OR ontologies OR meta-model* OR metamodel*) AND ("natural disaster" OR "natural disasters")) ». Seuls les articles catégorisés sous la dénomination « Computer Science » ont été retenus. En tout, 20 publications ont été obtenues, puis étudiées. Les résultats les plus pertinents sont résumés dans le Tableau 2.2.

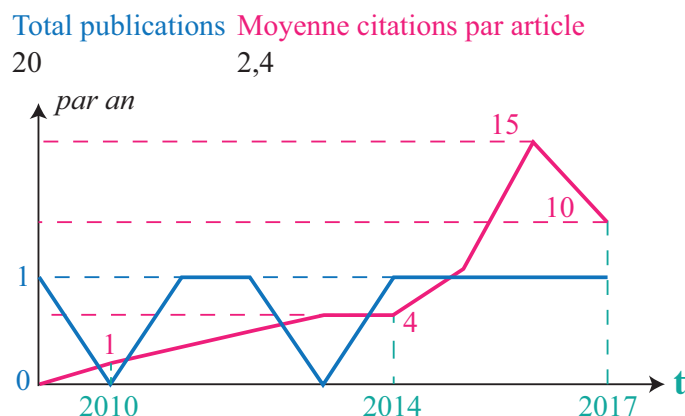


FIGURE 2.6 – Résultats obtenus, le 23 mai 2018, sur [Web of Science](#), sur la base des résultats des *Articles* publiés après 2008 et sous la catégorie *Computer Science*, à la requête : ((ontology OR ontologies OR meta-model* OR metamodel*) AND ("natural disaster" OR "natural disasters"))

La figure 2.6 présente l'évolution du nombre d'articles répondant à notre requête sur la base Web of Science. Cette évolution est nuancée. Pourtant, selon l'étude présentée en Annexe C, le nombre d'articles et le nombre de citations, portant sur les structures d'informations liées à une situation de crise, augmentent régulièrement depuis 1992.

2.3.3 La structuration des informations en situation de crise

Pour répondre aux besoins d'information des décideurs en gestion de crise, la structure de référence recherchée devrait permettre aux décideurs de :

2. Les étapes de modélisation d'une situation de crise dans la littérature

Réf.	Complexité	Couverture	Niveau d'abstraction	Concepts principaux
Bennani et al. (2017)	Ontologie	Objectif, Partenaire	Domaine (attaque terroriste)	Attaque evaluation, Context, Cause, Solution
Benaben et al. (2017)	Métamodèle	Partenaire, Contexte, Objectif	Mixte (collaboration - crise)	Environment component, Opportunity/Threat, Partner, Capacity Planning engine, Event ontology, Task ontology
Han et Xu (2015)	Ontologie	Partenaire, Objectif	Général	Response organisation, Rescue, Exposure
Othman et al. (2014)	Métamodèle	Partenaire, Contexte, Objectif	Général (réponse aux crises)	Disaster event, Cell, Infrastructure element, Agent
Kruchten et al. (2008)	Métamodèle	Partenaire, Contexte, Objectif	Général (crise)	
Jung et Chung (2015)	Ontologie	Partenaire, Contexte	Mixte	Environment, Location, Equipment, Personal
Calcaterra et al. (2015)	Ontologie	Contexte, Objectifs	Général (crise)	Hazard, Sensor, Spatial object, Region
Yu et al. (2015)	Ontologie	Contexte, Objectif, Partenaire	Domaine (réseau électrique)	Environment, Responders, Physical system, Hazard, Risk, Safety measure
Iribarne et al. (2010)	Métamodèle	Partenaire	Général	Actor, Coreography, Task
Zschocke et al. (2010)	Métamodèle	Objectif, Partner	Général (crise)	Hazard, Risk, Vulnerability, Actor, Corrective risk management

TABLE 2.2 – Travaux illustrés sur la Figure 2.5 et résultats pertinents de l'état de l'art sur les métamodèles de niveau 2 dédiés à la représentation de situation de crise, et notamment les crises dues à des catastrophes naturelles

- raisonner. Pour ce faire, les concepts, et par extension les instances du modèle, doivent être *liées* les unes aux autres grâce à des relations prédéfinies ;
- améliorer efficacement leur conscience de la situation. Pour ce faire, les concepts doivent *couvrir* l'ensemble des informations utilisables pour soutenir les prises de décisions. Ces informations peuvent concerner les partenaires de la collaboration de crise, leurs objectifs ou les caractéristiques du théâtre de crise ;
- suivre et réagir aux effets en cascade dus à la crise en cours. Pour ce faire, les concepts proposés doivent pouvoir se *spécifier* selon la nature changeante et imprévisible de la crise.

Tous les résultats présentés dans le Tableau 2.2, issus de l'état de l'art ou de la Figure 2.5, ont été évalués selon ces trois critères. Au vue de ces résultats, nous proposons deux catégories distinctes pour évaluer les structures existantes selon leur complexité :

- les métamodèles qui permettent de définir des relations complexes entre concepts, comme par exemple, « un risque, généré par un danger, impacte un ou plusieurs enjeux » (cf. Figure 1.3). Cette catégorie concerne les travaux de Benaben et al. (2017); Othman et al. (2014); Kruchten et al. (2008); Iribarne et al. (2010) et Zschocke et al. (2010).
- les métamodèles, que nous nommeront ontologies ici, qui sont mis au point pour définir et fixer des termes dans un domaine métier donné.

Puis trois catégories différentes selon leur couverture des informations portant sur une situation de crise :

- les *généralistes* qui concernent les objectifs de la collaboration, le contexte de crise et les partenaires impliqués dans la réponse à la crise. Nous retrouvons ici les travaux de Benaben et al. (2017); Othman et al. (2014); Kruchten et al. (2008) et Yu et al. (2015).
- les *intermédiaires*, qui concernent au moins deux de ces catégories ;
- les *spécialistes* qui conceptualisent les informations d'une seule de ces catégories, comme les travaux de Iribarne et al. (2010).

Enfin, trois catégories différentes selon le niveau d'abstraction travaillé :

- *général*, lorsque tous les types de crises peuvent être représentés par la structure proposée ;
- par *domaine*, lorsque les concepts proposés sont spécifiques à un type de crise, comme chez Bennani et al. (2017) ou Yu et al. (2015).
- *mixte*, lorsque la structure permet de décrire une crise selon plusieurs niveaux d'abstraction comme chez Benaben et al. (2017) et Jung et Chung (2015).

Dans notre contexte de crise, une seule structure de référence permet de répondre à nos trois critères : la structure proposée par Bénaben (2012) qui est complexe, généraliste et mixte.

2.4 Les méthodes pour interpréter les données collectées

Pour continuer à compléter notre cadre de réponse, cette partie s'intéresse à l'étape d'interprétation, en orange sur la Figure 2.2. Pour cette étape, l'objectif est de permettre au système d'information de se représenter l'environnement de crise courant de façon à générer des informations utilisables par les décideurs en cellule de crise. Pour ce faire, il doit modéliser la situation de crise selon le métamodèle choisi à la section 2.3.

2.4.1 Les types d'interprétation

Pour générer de l'information, le module d'interprétation peut faire appel, selon nous, à deux types de méthodes :

- les méthodes de *création* (c) qui permettent, à partir de données, d'instancier un métamodèle pour obtenir un modèle de la situation de crise ;
- les méthodes d'*amélioration* (a) qui permettent, à partir du modèle de situation, d'ajouter des instances ou de lier les instances entre-elles.

Ces deux types de méthodes participent au fonctionnement du système d'information recherché qui doit être capable de (i) créer le modèle d'une situation de crise à partir de données brutes, de (ii) le compléter en inférant de nouvelles informations de manière à couvrir l'ensemble des informations utiles aux décideurs, et (iii) de tenir à jour le modèle obtenu. Pour ce faire, ces deux méthodes utilisent l'une des techniques listées par [Perera et al. \(2014\)](#) :

- des techniques d'apprentissage, dont :
 - (c) l'*apprentissage supervisé* où un système met au point une fonction capable de catégoriser de nouvelles entrées, sur la base d'exemples fournis par des experts ;
 - (c) l'*apprentissage non supervisé* qui cherche à identifier des structures au sein d'un ensemble de données non expertisées.
- des règles, dont :
 - (c,a) les *règles conditionnelles* de type *Si, Alors, Sinon* utilisées pour raisonner sur un modèle de situation existant, pour détecter des événements plus ou moins complexes, ou pour annoter des données entrantes et les associer à des classes existantes ;
 - (a) les *règles d'inférences* utilisées sur des modèles de situation existants pour inférer de nouvelles informations sur la base des anciennes.
- l'un des principes formant :
 - (c,a) la *logique probabiliste* qui utilise les probabilités pour combiner des données, identifier des solutions, ou comprendre l'origine de certains événements ;
 - la *logique floue* où des affirmations plus nuancées que la dichotomie vrai/faux sont utilisées. Elle peut être utilisée en supplément d'une des méthodes ci-dessus.

La logique floue ne sera pas étudiée dans ce manuscrit : elle pourra toujours permettre, dans un second temps, d'améliorer la ou les méthodes qui seront choisies pour compléter le système d'information recherché.

2.4.2 Les méthodes d'interprétation dans la littérature

Pour classifier les principales méthodes d'interprétation utilisées aujourd'hui, nous avons interrogé la base documentaires **Web of Science** avec la dernière des requêtes décrites sur la Figure 2.3 : « (event* AND awareness AND (model* OR ontolog*) AND decision*) ». Seuls les articles catégorisés sous la dénomination « Computer Science » ont été retenus. En tout, 13 publications ont été obtenues, puis étudiées. Les résultats les plus pertinents sont résumés dans le Tableau 2.1.

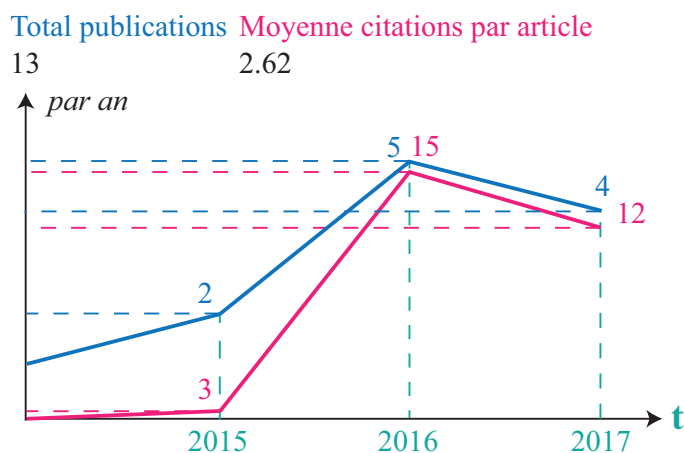


FIGURE 2.7 – Résultats obtenus, le 23 mai 2018, sur **Web of Science**, sur la base des *Articles* publiés après 2014 et sous la catégorie *Computer Science*, à la requête : (event* AND awareness AND (model* OR ontolog*) AND decision*)

La Figure 2.7 présente l'évolution du nombre de publications, répondant à notre requête sur la base Web of Science. Cette évolution est nuancée mais le faible nombre de publications observées ne permet pas de généraliser ces observations. Nous pouvons simplement noter que les premières publications coïncident avec (i) les premières utilisations des publications obtenues à la section 2.2 pour la détection d'évènements et (ii) à l'augmentation du nombre de publications parues sur la modélisation de situation de crise (cf. Annexe C).

2.4.3 L'interprétation de données en situation de crise

Le Tableau 2.3 présente les résultats de l'état de l'art présenté à la Figure 2.7. Nous proposons d'évaluer les types d'interprétation (cf. section 2.4.1), selon leur capacité à gérer les 4Vs du Big data lors de leur utilisation dans les travaux présentés. Ici, seules les problématiques liées à la véracité, au volume et à la vitesse sont prises en compte : la variété des données est éliminée par notre hypothèse 3 et la variété des informations est éliminée grâce à la mise en place d'un métamodèle.

La Figure 2.8 reprends les résultats présentés sur le tableau 2.3, regroupés selon leur utilisation des méthodes d'apprentissage, des règles conditionnelles, des règles

2. Les étapes de modélisation d'une situation de crise dans la littérature

Réf.	Type	Objet de recherche	Solution proposée	4Vs ?
Akbar et al. (2018)	(c) RC (a) P	Analyser en temps réel les données provenant de l'Internet des objets	Un moteur d'interprétation et des réseaux bayésiens pour informatiser les incertitudes	vitesse véra- cité
Taleb et al. (2018)	(c) RC, A	Optimiser le taux d'utilisation des données récoltées par le système, en activant/désactivant le module de collection	Des fenêtres de temps activées par des règles, basées sur un modèle de la situation courante	volume vélo- cité
Botega et al. (2017)	(a) RI	Évaluer la pertinence et la véracité des informations disponibles	Une analyse sémantique des informations textuelles disponibles	volume véra- cité
Pai et al. (2017)	(a) RI, RC	Déduire automatiquement de nouvelles informations sur la base des anciennes	Une ontologie adaptée à la situation, une analyse sémantique et un raisonnement à base de règles	variété
Basu et al. (2016)	(c,a) RC	Agréger des données provenant d'un grand nombre de capteurs de même nature	Un métamodèle et des règles pour détecter, puis agréger des événements selon une fenêtre de temps	vitesse variété
Jefferson et Johannes (2016)	(a) P	Modéliser une situation de crise alors que la collecte des données est incomplète	Des scénarios générés à partir de données incomplètes puis évaluer selon leur probabilité	vitesse
Soeanu et al. (2016)	(a) P	Adresser les flux d'instructions et d'informations au bon niveau hiérarchique	Une fenêtre de temps et la formalisation du processus de communication	volume
Wen et al. (2015)	(a) A	Reconnaître des activités dans un réseau de capteur donné	Des règles d'association entre un motif et une activité pour reconnaître les activités en cours	vitesse
Aloulou et al. (2015)	(a) RC	Améliorer la véracité des informations manipulées en intégrant les incertitudes liées aux capteurs	Une classe dédiée aux incertitudes (cf. métamodèle) et la propagation des incertitudes lors de l'interprétation des données	véracité

TABLE 2.3 – Résultats les plus pertinents de l'état de l'art sur les méthodes dédiées à (c) la création d'informations ou (a) l'amélioration d'un modèle de situation existant. Les solutions reposent sur des méthodes d'apprentissage (A), des règles conditionnelles (RC), de règles d'inférence ou sémantiques (RI) ou la logique probabiliste (P)

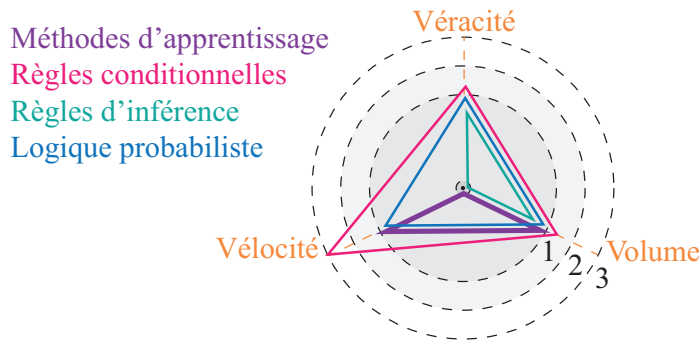


FIGURE 2.8 – Nombre de travaux, par type, de notre état de l’art proposant de traiter un des 4Vs (hors variété)

d’inférence ou de la logique probabiliste. Chaque groupe comptabilise ensuite le nombre de fois où la véracité, le volume ou la vélocité des informations a pu être contrôlée. D’après nos résultats, l’utilisation de règles permettrait, comme la logique probabiliste de contrer trois des 4Vs du Big Data :

- les règles conditionnelles ou d’inférence peuvent interpréter des événements en temps réel (Akbar et al., 2018; Taleb et al., 2018; Basu et al., 2016), en vue d’obtenir des informations sur la situation de crise en cours, en tenant compte des incertitudes (Aloulou et al., 2015) ;
- la logique probabiliste permet d’améliorer la véracité du modèle obtenu Akbar et al. (2018), déduire de nouvelles informations (Pai et al., 2017; Jefferson et Johannes, 2016) ou optimiser le taux d’utilisation des informations composant la COP (Soeanu et al., 2016).

Méthode unitaire	(c)	(a)	Types d’évènements en entrée	Types d’informations en sortie
Apprentissage supervisé	•		x	1
Apprentissage non supervisé	•		x	1
Règles conditionnelles	•	•	x	x
Règles d’inférence		•	x	x
Logique probabiliste	•	•	x	1

TABLE 2.4 – Résumé des techniques d’interprétations utilisées dans la littérature, qu’elle soit dédiée à la création (c) d’informations ou à l’amélioration (a) d’un modèle existant, à partir de 0, 1 ou (x) types de données pour instancier 0, 1 ou (x) types de concepts

Le Tableau 2.4 illustre les caractéristiques des méthodes d’interprétation trouvées dans nos différents états de l’art. Les avantages et inconvénients de toutes ces méthodes ont été listés par Perera et al. (2014). Nous retiendrons qu’aucun des travaux retrouvés ici ne propose d’utiliser la logique probabiliste pour *créer* de l’information à partir de données. Ce qui peut s’expliquer par le caractère imprévisible des crises compliquant l’utilisation de cas passés : les règles probabilistes sont utilisables seulement lorsque les probabilités sont connues.

Les résultats présentés sur la Figure 2.8 et le Tableau 2.4 nous pousse donc à rechercher une méthode d’interprétation utilisant des règles conditionnelles ou des règles d’inférence. Ces règles devront être rédigées à la main, pendant la phase de préparation, avant l’évènement majeur, et être validées par des experts métiers, habitués aux situations de crise.

2.5 Les solutions transversales

Pour finir, cette section tente d'identifier des méthodes transversales aux autres méthodes composant notre cadre de réponse (cf. Figure 2.2). Pour rappel, les sections précédentes nous ont permis de caractériser le système d'information recherché, pour qu'il soit adapté aux situations de crise. Il doit comprendre :

- une méthode de collecte de type (x à x) (cf. section 2.2) ;
- un métamodèle permettant (i) d'établir des relations complexes entre concepts (*complexe*), (ii) de décrire tous les aspects d'une situation de crise (*généraliste*) et (iii) de définir des concepts par type de crise ou pour tout type de crise (*configurable*) ;
- une méthode d'interprétation utilisant des règles conditionnelles ou des règles d'inférence.

2.5.1 Les méthodes transversales dans la littérature

Trois requêtes ont été utilisées pour identifier des méthodes transversales à nos besoins (cf. Figure 2.3). Elles combinent ensemble les trois requêtes utilisées aux sections 2.2, 2.3 et 2.4, et elles concernent des travaux publiés depuis dix ans par des journaux ou par certaines conférences. Les résultats sont représentés sur la Figure 2.9 et les plus pertinents, différents de ceux présentés aux états de l'art précédents, sont décrits dans le Tableau 2.5.

La Figure 2.9 présente l'évolution du nombre de publications répondant à nos trois requêtes sur la base Web of Science. Les outils, permettant de détecter des événements pour les interpréter en informations et améliorer la conscience de la situation de l'utilisateur, semblent très recherchés depuis 2013 : le nombre de citations par article paru ne cesse d'augmenter. L'évolution de cette courbe, en orange, est similaire aux évolutions du nombre d'articles portant sur la conscience d'une situation de plus en plus cités depuis 2013¹. La courbe violette est, quant à elle, révélatrice du peu de publications essayant, comme nous le proposons, d'améliorer la conscience de la situation d'un agent (logiciel ou humain) à partir d'un ensemble d'informations structuré sous forme de modèle.

2.5.2 Les solutions transversales en situation de crise

Le Tableau 2.5 présente les résultats de l'état de l'art présenté à la Figure 2.9. Au vu des résultats, nous proposons d'évaluer les 72 publications obtenues selon leur capacité à couvrir nos trois besoins.

Pour commencer, nous cherchons une méthode de collecte de type (x à x) pour caractériser une situation dans son ensemble et non pas une sous-partie caractérisée par un seul objet d'intérêt comme un bâtiment (Brizzi et al., 2016), un patient (Gómez et al., 2016) ou une pièce (Evchina et Lastra, 2016). Aucun des travaux présentés (cf. Tableau 2.5) ne permet de récolter x types de données pour en déduire x types d'informations. Par contre, il est envisageable de combiner des

1. En 2013, Web of Science comptait 3265 citations des 4181 articles parus depuis 2008 pour la requête ("context awareness" OR "contextual awareness" OR "situation awareness" OR "situational awareness"). En 2017, Web of Science comptait 10607 citations des 8795 parus depuis 2008.

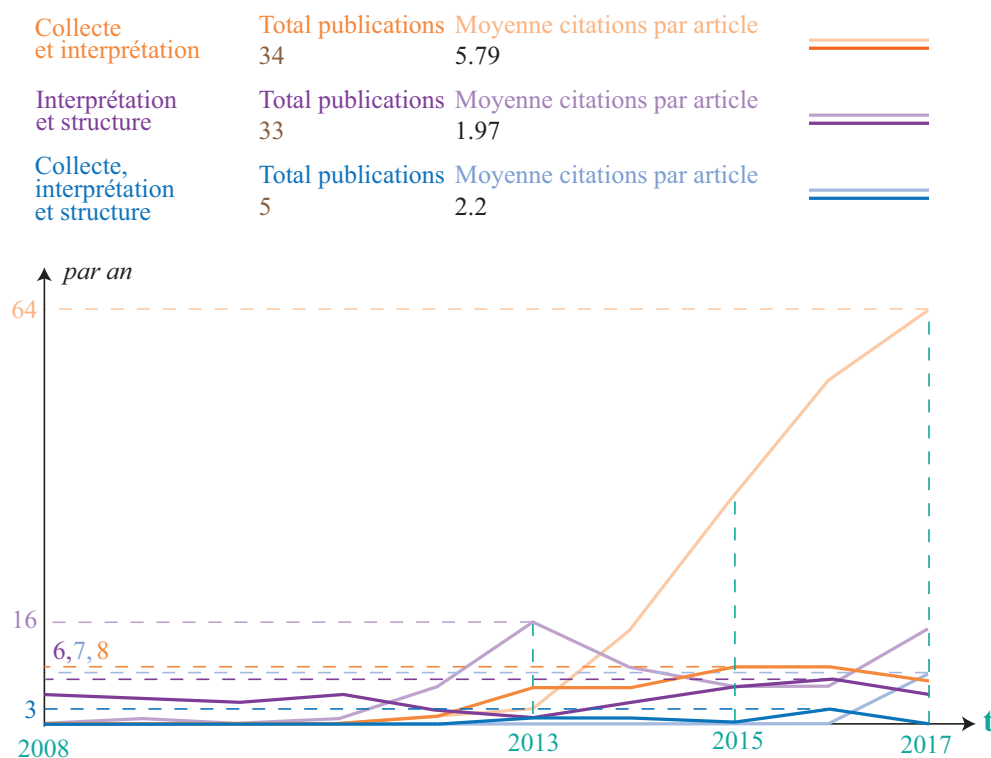


FIGURE 2.9 – Résultats obtenus, le 12 juin 2018, sur [Web of Science](#), sur la base des articles de journaux ou de conférences publiés après 2008 et sous la catégorie *Computer Science*, aux requêtes : *(("social media" OR microblog* OR "internet of things" OR "volunteered geographic information") AND real-time AND ("context awareness" OR "contextual awareness" OR "situation awareness" OR "situational awareness") AND information AND (data OR event)); (event* AND awareness AND information AND (ontology OR ontologies OR meta-model* OR metamodel*)); et (("social media" OR microblog* OR "internet of things" OR "volunteered geographic information") AND real-time AND awareness AND information AND (ontology OR ontologies OR meta-model* OR metamodel*))*

2. Les étapes de modélisation d'une situation de crise dans la littérature

Réf.	Quoi ?	Collecte ?	Interprétation ?	Structure ?
Alam et al. (2018)	Catégoriser les images des réseaux sociaux	(1 à x)	(c) Apprentissage par classifieurs personnalisables	Classification
Le Sceller et al. (2017)	Détecter les cyberattaques via Twitter	(1 à x)	(c) Apprentissage par clusters sémantiques, puis géographiques	Taxonomie générale et par type de cyber-attaque.
Lechner et al. (2017)	Détecter des événements sur une chaîne de production	(1 à x)	(c) Règles conditionnelles	Typologie des événements à détecter
Kim et al. (2016)	Détecter les risques sur une chaîne du froid, via la RFID	(1 à x)	(c+a) Règles d'inférence et un agent EPA	Ontologie pour une partie des informations
Evchina et Lastra (2016)	Interpréter les flux de données d'un système de surveillance	(? à 1)	(c+a) Règles d'inférence et moteur CEP	Métamodèle autour d'un lieu (pièce) à identifier
Andrews et al. (2016)	Identifier des groupes de tweets	(1 à x)	(c+a) Règles et Agrégation par attributs similaires	Taxonomie par type de crise et d'incident
Kabir et al. (2016)	Modéliser une situation d'après l'agenda d'un utilisateur	Absente	(a) Règles d'inférence	Ontologie des types d'activité et de période
Gómez et al. (2016)	Surveiller à distance l'état d'un patient	(x à 1)	(c) Règles seuils pour générer des alertes	Ontologie par type d'utilisateur
Brizzi et al. (2016)	Modéliser une ville et ses réseaux d'énergie	(x à ?)	Manuelle via un tableau de bord	Métamodèle autour d'un bâtiment
Steed et al. (2015)	Aider la détection des tendances sur les réseaux sociaux	(1 à x)	(c) Apprentissage grâce à un classifieur	Classification de 3 types d'émotion
Lee et al. (2014); Lee (2012)	Identifier des événements sur les réseaux sociaux	(1 à x)	(c) Apprentissage par clusters géographiques et équations mathématiques	Aucune structure de référence

TABLE 2.5 – Résultats les plus pertinents de l'état de l'art sur les méthodes dédiées à la fois à la collecte de données et à leur interprétation (création (c) ou amélioration (a)) en un ensemble d'information modélisant une situation donnée

méthodes (1 à x) comme celles proposées par [Alam et al. \(2018\)](#); [Le Sceller et al. \(2017\)](#); [Lechner et al. \(2017\)](#); [Kim et al. \(2016\)](#); [Andrews et al. \(2016\)](#); [Steed et al. \(2015\)](#) ou [Lee et al. \(2014\)](#) pour décrire l'ensemble d'une situation observée.

Deuxièmement, nous cherchons un métamodèle complexe, généraliste et configurable. Ici, le métamodèle le plus généraliste est proposé par [Kim et al. \(2016\)](#) qui cherchent à couvrir tous les aspects d'une chaîne du froid, mais il ne semble ni complexe, ni configurable. Un métamodèle configurable est proposé par [Evchina et Lastra \(2016\)](#) : il peut s'adapter au type de capteurs ou au type de pièces à représenter, par contre, il n'est pas complexe. Les seuls travaux présentés ici qui ont cherché à définir des relations complexes entre concepts sont présentés par [Brizzi et al. \(2016\)](#), mais leur métamodèle n'est ni généraliste, ni configurable.

Pour finir, nous cherchions une méthode d'interprétation capable de gérer la vélocité, le volume et la véracité des données brutes récoltées :

- seuls les travaux d'[Evchina et Lastra \(2016\)](#) semblent ici capable de suivre la *vélocité* des données entrantes : selon lui, l'utilisation d'un moteur CEP permet un temps d'interprétation, à l'étape de création (cf. section 2.4), de l'ordre de la seconde. Il a ainsi connecté plus de 10 000 capteurs à son système et il est parvenu à suivre un rythme de 1000 mises à jour par seconde.
- le *volume* et la valeur des informations obtenues sont principalement contrôlés par l'utilisation d'une structure de référence aboutie, comme chez [Kim et al. \(2016\)](#); [Evchina et Lastra \(2016\)](#) ou [Brizzi et al. \(2016\)](#). Les travaux utilisant des méthodes de collecte (x à 1) ou (x à x), plus sujet à des volumes de données importants, utilisent quant à eux un ensemble de règles fini, comme c'est le cas chez ([Gómez et al., 2016](#)).
- la recherche d'une meilleure véracité des informations obtenues est, quant à elle, au centre des méthodes basées sur de l'apprentissage, comme chez [Alam et al. \(2018\)](#); [Le Sceller et al. \(2017\)](#); [Steed et al. \(2015\)](#) ou [Lee et al. \(2014\)](#).

En résumé, parmi ces 72 travaux, seuls les travaux présentés respectivement par [Kim et al. \(2016\)](#), et [Akbar et al. \(2018\)](#) proposent une méthode hybride. Les premiers sont capables d'interpréter des informations de manière structurées et les seconds sont capables de collecter des données et de les interpréter grâce à une même technologie.

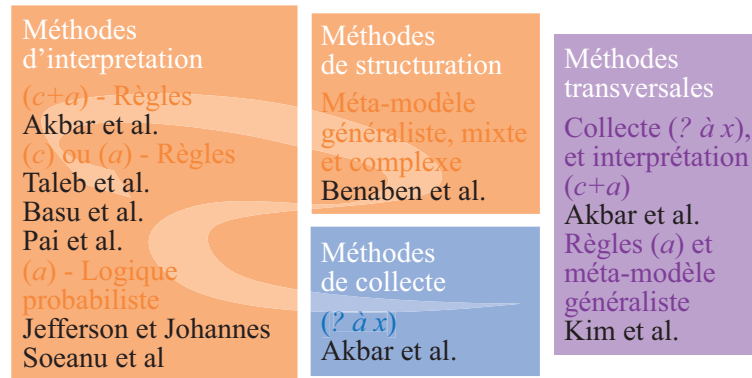
2.6 Récapitulatif des méthodes identifiées

Nous cherchions un système d'information capable d'alimenter en informations le niveau 2 de conscience des décideurs en cellule de crise.

La Figure 2.10 présente les travaux identifiés dans ce chapitre, aux différentes étapes de travail présentées sur la Figure 2.1. Les méthodes retenues permettent tour à tour de (i) détecter des évènements pour décrire multiples aspects d'une situation donnée ([Akbar et al., 2018](#)) (ii) d'interpréter des évènements pour créer de l'information ou améliorer un ensemble d'information existant ([Akbar et al., 2018](#); [Taleb et al., 2018](#); [Basu et al., 2016](#); [Pai et al., 2017](#); [Jefferson et Johannes,](#)

2. Les étapes de modélisation d'une situation de crise dans la littérature

Les méthodes identifiées, certaines transversales, adaptées à notre contexte de gestion de crise, soumis aux 4Vs du Big data



Abbréviations: (? à x) collecte d'un ou plusieurs types de données décrivant plusieurs types d'objets, (c) méthode de création d'instances, (a) méthodes d'amélioration d'un modèle existant.

FIGURE 2.10 – Réponse apportée au besoin d'informations des cellules de crise, nourrie par trois états de l'art et encadrée par quatre hypothèses de travail

2016; Soeanu et al., 2016), ou (iii) de structurer l'ensemble des informations disponibles grâce un un métamodèle généraliste, mixte et complexe (Benaben et al., 2017). Notre contribution se focalisera ici sur une méthode *transversale* capable de récupérer des données de différentes sources, pour générer, améliorer et mettre à jour en temps réel un modèle représentatif de la situation de crise. Dans ce cadre, la meilleure solution semble être de partir des méthodes utilisées par Akbar et al. (2018); Kim et al. (2016) et Benaben et al. (2017) :

- un moteur CEP, comme proposé par Kim et al. (2016) et Akbar et al. (2018) pour (i) récolter des événements selon une méthode de type (x à x) et (ii) utiliser des règles conditionnelles et des règles d'inférence efficaces pour contrer la *véracité*, la *vélocité* et le *volume* des données et informations manipulées ;
- un métamodèle hérité du métamodèle de Benaben et al. (2017) dédié aux collaborations de crise, configurable selon la nature de la crise pour structurer et formaliser l'ensemble des informations (*variées*) disponibles ;

Le chapitre 3 présente ainsi l'architecture proposée qui permettra de collecter et d'interpréter des données en un ensemble d'informations formalisées, dans un contexte de gestion de crise et de Big Data.