



HAL
open science

Représentations de connaissances spatiales évolutives : des ontologies aux géovisualisations

Marlène Villanova-Oliver

► **To cite this version:**

Marlène Villanova-Oliver. Représentations de connaissances spatiales évolutives : des ontologies aux géovisualisations. Intelligence artificielle [cs.AI]. Communauté Université Grenoble Alpes, 2018. tel-01935490

HAL Id: tel-01935490

<https://hal.science/tel-01935490v1>

Submitted on 26 Nov 2018

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.

Habilitation à Diriger des Recherches

Spécialité : Informatique

Ecole Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique

présentée par

Marlène Villanova-Oliver

Représentations de connaissances spatiales évolutives : des ontologies aux géovisualisations

Présentée publiquement le 20 Novembre 2018

Composition du Jury

Rapporteurs

Thérèse Libourel, Professeur des Universités, Eméritat,
Université de Montpellier

Christophe Claramunt, Professeur des Universités
Institut de Recherche de l'Ecole Navale de Brest (IRENav)

Roland Billen, Professeur ordinaire
Université de Liège, Belgique

Examineurs

Anne Ruas, Ingénieur en Chef des Ponts, Eaux et Forêts - HDR,
*Institut Français des Sciences et Technologies des Transports,
de l'Aménagement et des Réseaux (IFSTTAR)*

Eric Gaussier, Professeur des Universités,
Université Grenoble Alpes

Jérôme Gensel, Professeur des Universités,
Université Grenoble Alpes



R

emerciements

Je tiens à remercier très sincèrement mes trois rapporteurs pour avoir accepté d'évaluer mes travaux et pour avoir pris le temps nécessaire à la réalisation de cette tâche, dans toutes ses facettes : le casse-tête de la planification d'une date, la lecture du manuscrit, les questionnements suscités, la rédaction du rapport, le voyage jusqu'à Grenoble. Madame Thérèse Libourel, Monsieur Christophe Claramunt et Monsieur Roland Billen, je suis vraiment très honorée que vous ayez répondu favorablement à ma sollicitation. Je vous fais part de toute ma gratitude pour l'intérêt témoigné à mes recherches, pour les avis scientifiques que vous avez émis et pour les remarques constructives dont vous m'avez fait part.

Mes remerciements vont ensuite à Madame Anne Ruas et Monsieur Eric Gaussier. Vous m'avez fait l'honneur de votre participation à ce jury, et je dois dire que la spontanéité avec laquelle vous avez accepté m'a vraiment touchée. Ce jury, enfin, est complété par Monsieur Jérôme Gensel qui est pour beaucoup dans la réalisation de cette HDR, et au-delà, par tout ce qu'il m'a transmis, dans ma façon de mener à bien mes missions au quotidien.

Merci à vous six, j'ai beaucoup de respect pour vos travaux, votre expertise et vos qualités humaines. Je vous suis très reconnaissante de me donner l'occasion d'un échange scientifique avec vous en ce 20 novembre 2018.

J'ai croisé beaucoup de chemins tout au long de ce parcours qui m'amène aujourd'hui à rédiger cette page de remerciements. La dernière que j'ai écrite dans un contexte universitaire, en 2002, se terminait par « *Il y a un peu de chacun de vous dans cette thèse* ». Sur les plans professionnel et personnel, c'est un très grand bonheur que nombre d'entre ceux et celles auxquels je pensais alors soient encore autour de moi aujourd'hui. C'est aussi beaucoup de satisfactions tirées de toutes les nouvelles rencontres qui ont eu lieu. J'ai évidemment, aussi, une pensée toute particulière pour ceux qui nous ont quittés depuis...

Il y a un peu de chacun de vous dans cette HDR

Il y a un peu (beaucoup, en fait !) de chacun des membres de l'équipe Steamer au sein de laquelle j'ai un réel plaisir à évoluer depuis sa création en 2007. Merci à tous pour m'avoir témoigné confiance et soutien en me proposant d'en assumer la responsabilité à la suite de Jérôme et Paule-Annick. Quel précieux héritage est le mien ! Merci à Danielle, Philippe et Sylvain, qui complètent le cercle restreint des permanents de l'équipe. J'associe aussi Sonia, invitée permanente, à ces remerciements. Pouvoir compter sur vous à tout instant pour faire vivre cette équipe scientifiquement, techniquement et humainement est une chance. Steamer c'est aussi nombre de doctorants, d'étudiants

de master, d'ingénieurs ou encore de stagiaires qui, par leur créativité, leur curiosité, leurs questions ont fait avancer mes recherches. Mes plus sincères remerciements vont à chacun de vous, mais je me dois de faire une dédicace spéciale à « mes » doctorants : Marius, Dia, Anthony, Cécile, David, Camille, Lina, Matthieu. J'ai aussi appris beaucoup avec et grâce à vous.

Il y a un peu de chacun des membres du LIG avec qui je suis amenée à travailler, chercheurs et personnels administratifs et techniques. Nous sommes bien nombreux dans ce grand laboratoire et beaucoup d'entre vous ont, à un moment ou un autre, peut-être même sans le savoir, été une source d'inspiration, un exemple, un soutien ou même simplement là pour un moment d'échange et de partage. Merci à vous.

Il y a un peu de chacun de mes collègues du département STID de l'IUT2. Je vous remercie pour l'état d'esprit qui règne au sein de notre petite équipe, qui rend l'investissement en enseignement si naturel et agréable. C'est un plaisir d'œuvrer à vos côtés pour nos étudiants.

Il y a un peu de chacun des chercheurs et ingénieurs avec lesquels des collaborations se sont développées à diverses occasions : lors des événements du GDR MAGIS notamment, dans le cadre des projets (ANR, européens, autres) impliquant Steamer, ou encore pour des co-encadrements de thèses. Je suis reconnaissante pour cette richesse et cette ouverture scientifique que vous m'apportez.

Il y a un peu de chacun de mes amis et des membres de ma famille. Vous savez, je l'espère, combien vous comptez et ce que je dois à chacun de vous.... mais je ne manquerai pas de vous le redire autour du verre que je vous ai promis « quand tout ça serait bouclé ». J'ai hâte de vous retrouver !

Il y a tant de toi, Loïc, de toi, Evan et de toi, Marylou. Evidemment.

Pauline

Tables des Matières

Préambule	1
Les travaux post-thèse sur l'adaptation	1
L'ouverture vers le Web Sémantique Géospatial.....	2
Approches orientées ontologies pour la représentation de connaissances	3
Introduction	5
Représentation de connaissances dans le Web des Données Ouvertes et Liées	7
Géovisualisation	8
Organisation du Document	9
2. Découpages territoriaux évolutifs.....	11
2.1 Contexte des travaux.....	11
2.1.1 Motivations et enjeux.....	11
2.1.2 Ontologies pour la structuration d'un territoire et ses évolutions	12
2.1.3 Positionnement de l'approche	15
2.2 Représenter la structure du territoire.....	17
2.2.1 L'ontologie TSN	17
2.2.2 Cas d'étude : Description des composants de la NUTS.....	18
2.3 Représenter l'évolution d'un territoire	20
2.3.1 L'ontologie TSN-Change.....	20
2.3.1.1 Typologie des changements.....	20
2.3.1.2 Importance de l'Identité dans la détermination des changements	21
2.3.1.3 Principes pour l'expression des changements.....	21
2.3.1.4 Expression multi-niveau des changements.....	23
2.3.1.5 Gestion de l'évolution des frontières d'un territoire d'étude.....	24
2.3.2 Cas d'étude : Description des changements dans la NUTS	25
2.4 Algorithme de détection des changements entre versions de TSN	27
2.4.1 Identifier des changements entre versions de territoires	28
2.4.2 Regrouper et qualifier les changements.....	30
2.4.3 Relier les changements	30
2.4.4 Produire le graphe RDF.....	30
2.4.5 Cas d'étude : détection de changements dans la NUTS.....	31
2.5 Exemple d'extractions de connaissances	32
2.5.1 Requêtage du SPARQL Endpoint.....	32
2.5.2 Ouverture sur le Web des LOD	33
2.6 Synthèse et prolongements.....	34

3. Trajectoires de vie personnelles.....	37
3.1 Contexte des travaux.....	37
3.1.1 Motivations et Enjeux.....	37
3.1.2 Modélisation des trajectoires de vie et des choix associés.....	39
3.1.3 Positionnement de l'approche.....	40
3.2 Concepts clés pour la représentation des trajectoires de vie.....	41
3.2.1 Structurer des trajectoires de vie multidimensionnelles.....	41
3.2.2 La granularité de l'information considérée.....	43
3.2.3 Lier l'approche multidimensionnelle et les facteurs explicatifs.....	44
3.3 Modélisation de trajectoires de vie : l'approche LTOP.....	45
3.3.1 Le patron de création d'une trajectoire thématique.....	46
3.3.1.1 Cas général.....	46
3.3.1.2 Trajectoire thématique à composante géographique ou métaphorique.....	47
3.3.2 Le patron d'expression de facteurs explicatifs.....	48
3.3.2.1 Cas général.....	48
3.3.2.2 Distinction entre <i>EventFactor</i> et <i>StateFactor</i>	48
3.3.2.3 <i>Facteurs</i> explicatifs externes à la trajectoire personnelle.....	50
3.4 Framework dédié à la gestion de trajectoires de vie.....	50
3.4.1 Création du modèle d'ontologie d'application.....	51
3.4.2 Acquisition de données.....	52
3.4.3 Enrichissement des données collectées.....	54
3.4.4 Exploitation des données de trajectoires.....	55
3.5 Synthèse et prolongements.....	56
4. Récits de situations de crise.....	59
4.1 Contexte des travaux.....	59
4.1.1 Motivations et Enjeux.....	59
4.1.2 Représentations des risques naturels et de leurs effets.....	61
4.1.3 Positionnement de l'approche.....	62
4.2 Ontologie d'événements issus de risques et leurs effets dominos.....	63
4.2.1 Vue d'ensemble sur l'ontologie IDISFER.....	63
4.2.2 Aperçu des différentes parties de l'ontologie.....	64
4.2.2.1 Ontologie relative au système ferroviaire.....	64
4.2.2.2 Ontologie relative aux phénomènes d'inondation.....	65
4.2.2.3 Ontologie relative aux incidents.....	65
4.2.2.4 Ontologie relative aux perturbations.....	65
4.2.2.5 Ontologie relative aux réactions.....	66
4.2.3 Exploiter l'ontologie IDISFER pour représenter les situations passées.....	66
4.3 De l'ontologie à la géovisualisation.....	68
4.3.1 Contexte, travaux préliminaires et positionnement d'une nouvelle approche...	68
4.3.2 Prérequis pour une vision systémique des impacts des inondations.....	70
4.3.3 Homogénéisation des granularités spatiales et temporelles.....	70
4.3.4 Représentation du type d'événement.....	72
4.3.5 Représentation de la date des événements.....	74
4.3.6 Visualisation d'événements de même localisation.....	78
4.3.7 Visualisation des relations de causalité entre événements.....	79
4.3.8 Visualisation des relations de causalités lointaines.....	80
4.3.9 Visualisation des temporalités.....	81
4.4 Prototype de géovisualisation dynamique du récit des événements.....	83
4.4.1 Vue d'ensemble.....	83
4.4.2 Animation du récit de situation de crise.....	84
4.4.3 Améliorer la représentation des effets dominos.....	86
4.5 Synthèse et prolongements.....	87

5. Conclusion.....	91
5.1 Bilan	91
5.2 Perspectives de Recherches.....	93
5.2.1 Cadrage et Enjeux.....	93
5.2.2 Des stratégies de raisonnement à formaliser.....	94
5.2.3 Un raisonnement étayé et traçable.....	95
5.2.4 Une sémiologie utile, utilisable et non intrusive.....	98
Annexes	99
Références bibliographiques	105
Encadrement Doctoral.....	115

Tables des Illustrations

Figures

Figure 1 - Association de deux approches pour les ontologies TSN et TSN-Change.....	16
Figure 2 - Principales classes et prédicats associés de TSN Ontology	17
Figure 3 - Historique de la NUTS	18
Figure 4 - Principales classes de l'ontologie TSN-Change	20
Figure 5 - Division d'une unité : distinction entre Scission et Extraction	21
Figure 6 - Illustration de changements décrits à l'aide de TSN-Change Ontology	22
Figure 7 - Représentation d'un changement dans TSN-Change Ontology.....	23
Figure 8 - Graphe multi-niveaux simplifié d'une fusion de deux territoires fictifs	24
Figure 9 - Graphe multi-niveaux simplifié de la scission de l'unité territoriale ES63.....	26
Figure 10 - Changements subis par l'unité L2_ES63 de la NUTS 1999 à la NUTS 2010	35
Figure 11 - Exemple de visualisation cartographique de changements territoriaux	36
Figure 12 - Extrait d'une trajectoire de vie.	42
Figure 13 - Représentation simplifiée du <i>Thematic Trajectory Pattern</i>	46
Figure 14 – Extensions pour les trajectoires à composante géographique et métaphorique	47
Figure 15 - Représentation schématique du <i>Trajectory Explanatory Factors Pattern</i>	48
Figure 16 - Illustration simplifiée d'un facteur explicatif de type <i>EventFactor</i>	49
Figure 17 - Illustration simplifiée d'un facteur explicatif de type <i>StateFactor</i>	49
Figure 18 - Architecture pour la représentation, la collecte et l'analyse de trajectoires de vie	51
Figure 19 - Représentation de étapes de la démarche de modélisation	51
Figure 20 - Exemple d'interface développée pour la collecte de données des trajectoires résidentielles.....	53
Figure 21 - Modèle simplifié de l'ontologie IDISFER	64
Figure 22 - Illustration simplifiée du concept <i>InCrisisEvent</i> dans IDISFER.....	66
Figure 23 - Illustration d'une instanciation de l'ontologie IDISFER sur l'exemple de l'inondation de 1940 dans les Pyrénées-Orientales.....	67
Figure 24 - Schématisation de l'approche <i>GVR Matching</i>	69
Figure 25 - Méthode de simplification des événements de granularités spatiales et temporelles hétérogènes.....	71
Figure 26 - Exemple de dérivation de concepts pour construire une classification exploitée dans la légende.....	73
Figure 27 - Symboles affectés aux classes d'événements dérivées à partir de l'ontologie IDISFER.....	74
Figure 28 - Reproduction du Cercle chromatique Ewald Hering.....	76
Figure 29 - Illustration de l'instanciation du concept <i>DateInLegend</i> de l'ontologie <i>GVR Matching</i>	77
Figure 30 - Exemples de légendes rendant compte des périodes de la situation de crise	77
Figure 31 - Illustration de la méthode de la pile chronologique	78
Figure 32 - Affichage à la demande des relations de causes et conséquences d'un événement.	80
Figure 33 - Affichage d'une relation de causalité distante.....	81
Figure 34 - Ligne à temps régulier montrant les événements survenus suite à l'inondation de 1947 à Metz.	82
Figure 35 - Ligne à temps proportionnel montrant les événements survenus suite à l'inondation de 1947 à Metz.	82
Figure 36 - Prototype de géovisualisation dédié au récit d'une situation de crise – Vue 1	84
Figure 37 - Prototype de géovisualisation dédié au récit d'une situation de crise – Vue 2	84

Figure 38 - Illustration de l'animation de la carte pour rendre compte des relations entre événements.....	85
Figure 39 - Exemple d'affichage de l'ensemble de la chaîne de causalité menant à un événement et en partant .	87
Figure 40 - Exemple d'un graphique en réseau pour chaîne de causalité menant à un événement et en partant ..	87
Figure 41 - Prototypage d'une interface d'acquisition d'indices avec imperfection	97
Figure 42 – Illustration de la désactivation d'un indice pour ne plus en tenir compte dans le raisonnement.....	97
Figure 43 - Vue générale de l'infrastructure logicielle STedi	101

Extraits de Codes

Extrait de code 1 - Description automatiquement générée d'une unité territoriale selon l'ontologie TSN	19
Extrait de code 2 - Patrons de construction des URIs des éléments d'une TSN	20
Extrait de code 3 - Vision simplifiée de l'algorithme de détection des changements entre versions de TSN.....	28
Extrait de code 4 - Requête SPARQL pour extraire les points d'entrées de graphes multi-niveaux du changement	32
Extrait de code 5 - Extrait du résultat retourné par la requête SPARQL présentée dans l'Extrait de code 4	32
Extrait de code 6 - Requêtes SPARQL DESCRIBE pour obtenir la description d'un nœud.....	32
Extrait de code 7 – Extrait de la chaîne des changements de l'UT ES63 entre les versions NUTS 1999 et 2003 33	

Equation

Équation 1 – Calcul de similarité entre deux unités appartenant à des versions différentes de TSN	29
---	----

Tableau

Tableau 1 – Visées descriptive et explicative des connaissances représentées dans les trois approches présentées	92
--	----

« *Le commencement de toutes les sciences,
c'est l'étonnement de ce que les choses sont ce qu'elles sont.* »

Aristote (384 av. J.-C. - 322 av. J.-C.)

Préambule

Soutenue en 2002, ma thèse a porté sur la définition, la formalisation et la mise en œuvre dans divers types de Systèmes d'Information de la notion d'accès progressif qui permet à un utilisateur d'accéder graduellement à un ensemble d'informations adapté à son profil, évitant ainsi la surcharge cognitive. Ces premiers travaux sur l'adaptabilité, que j'ai menés au sein du LSR¹ puis du LIG, ont été étendus dans différentes directions dans le cadre de thèses ou de masters auxquels j'ai collaboré. Je présente dans un premier temps, succinctement, ces travaux qui ont constitué mes axes principaux de recherche post-thèse (jusqu'en 2007 environ) et dont le point commun est la notion d'adaptation. J'évoque ensuite comment mes recherches se sont orientées vers le domaine du Web Sémantique et les approches de représentation de connaissances à base d'ontologies, cadres scientifiques dans lesquels s'inscrivent la plupart de mes travaux. Ce préambule est aussi l'occasion pour moi de présenter rapidement les trois premières thèses pour lesquelles j'ai officiellement tenu un rôle d'encadrante. La liste des DEA, masters et thèses que j'ai encadrés est donnée page 115.

Les travaux post-thèse sur l'adaptation

Dans la continuité de ma thèse, des travaux visant à poursuivre l'étude des possibilités d'expression de l'adaptation ont été engagés au sein de l'équipe. Dans la thèse de M. Kirsch-Pinheiro (soutenue en septembre 2006, (Kirsch Pinheiro, 2006)) sur l'adaptabilité et la conscience de groupe dans les Systèmes d'Information collaboratifs, nous avons travaillé à la définition d'un modèle de contexte exploité pour diffuser l'information adéquate à l'utilisateur. Avec A. Carrillo-Ramos (thèse soutenue en mars 2007, (Carrillo Ramos, 2007)), nous avons étudié la représentation des préférences des utilisateurs et les moyens d'en tenir compte selon le contexte courant d'utilisation. Dans la continuité de ces travaux, la thèse de Céline Lopez-Velasco (soutenue en novembre 2008, (Lopez-Velasco, 2008)) nous a permis, par une extension du langage WSDL², de proposer la définition de services Web capables de s'adapter à la fois aux caractéristiques de l'utilisateur (en termes de besoins, de disponibilité, de localisation, de préférences, *etc.*), et, aux caractéristiques du contexte d'utilisation (capacités du poste client, bande passante, *etc.*). Ces travaux auxquels j'ai collaboré m'ont permis de construire une vision encore plus large de l'apport, des enjeux et des défis liés à l'adaptation dans la

¹ Laboratoire Logiciels, Systèmes, Réseaux (UMR 5526) dont les équipes ont ensuite intégré en 2007 le Laboratoire d'Informatique de Grenoble (UMR 5217).

² Web Services Description Language, <https://www.w3.org/TR/2007/REC-wsdl20-20070626/>

conception et la mise en œuvre des Systèmes d'Information telle que je l'avais abordée dans ma thèse. Cette vision a également été alimentée en 2002/2003 dans le cadre du DEA de Ioan-Marius Bilasco qui a porté sur la conception et l'implantation de modèles permettant de spécifier des présentations multimédias adaptables. Il s'agissait d'étendre les capacités des feuilles de style et standards de présentation multimédias existants (comme SMIL³) afin d'adapter aux profils des utilisateurs la présentation des pages Web générées dynamiquement par un Système d'Information multimédia. Ces travaux ont été prolongés dans le cadre de la thèse de I-M Bilasco [TH-1]⁴ (soutenue en décembre 2007, (Bilasco, 2007)) et intitulée « *Une approche sémantique pour la réutilisation et l'adaptation de données 3D* ». Les recherches ont porté sur la modélisation et l'indexation sémantique de scènes 3D afin de permettre l'interrogation et la construction de nouvelles scènes par réutilisation de composants 3D. Au cours de cette thèse, l'intérêt des approches sémantiques pour l'adaptation s'est clairement imposé. D'autres recherches, menées en parallèle, confortaient l'idée que les approches sémantiques pouvaient notamment permettre des avancées dans l'exploitation des données relatives à l'espace et au temps.

L'ouverture vers le Web Sémantique Géospatial

Fin 2005, je me suis plus particulièrement intéressée au domaine de la représentation de connaissances par objets, et, en particulier, à son apport pour le Web Sémantique, dans le cadre du Master 2 Recherche de Alina Dia Miron. Ce travail visait à étudier les conditions du rapprochement d'un système de représentation par objets et du standard de description d'ontologies OWL⁵. Ces travaux se sont poursuivis par la thèse de A. D. Miron [TH-2]. Cette thèse intitulée '*Représentation de connaissances par objets pour le Web Sémantique Géospatial*' a été soutenue en décembre 2009 (Miron, 2009). La motivation à l'origine de ce travail était qu'une certaine catégorie d'informations présentes sur le Web était alors inexploitée : les informations relevant de l'espace et du temps. Si plus de 80% des pages Web contenaient (et contiennent encore) des références spatiales et temporelles (noms de lieux, adresses, coordonnées géographiques, dates, locutions temporelles, *etc.*), ces dernières n'étaient pas utilisées par les moteurs de recherche pour préciser le contexte d'une requête, pour la désambiguïsation de celle-ci, pour la classification des résultats, *etc.* Partant de ce constat, nous nous sommes intéressés aux techniques de représentation et de raisonnement à base d'annotations spatiales et temporelles, indispensables à l'émergence d'un Web Sémantique dit *Géospatial*, c'est-à-dire capable, par extension, de gérer les dimensions spatiale et temporelle de l'information. L'objectif du Web Sémantique Géospatial est, en ce sens, identique à celui du Web Sémantique : associer aux données spatio-temporelles des descriptions (métadonnées) interprétables par les humains, mais surtout par programmes à des fins d'automatisation. Un raisonneur spatial et temporel, compatible avec le langage standard de définition d'ontologies OWL a donc été mis au point dans le cadre de la thèse de A. D. Miron. Le système, appelé ONTOAST, est capable d'exploiter à la fois des données spatiales et temporelles quantitatives (*i.e.* les coordonnées géométriques des objets spatiaux, les intervalles de temps ou/et les instants) et des relations spatiales et temporelles qualitatives pour déduire des relations spatiales et temporelles jusque-là implicites. Un des prolongements donnés à ce travail a été l'étude d'une nouvelle technique de fouille du Web Sémantique, appelé analyse sémantique, qui vise la découverte des relations directes et indirectes qui existent entre deux individus. Ce paradigme de recherche permet ainsi de répondre à des questions du type : « l'instance x est-elle liée, d'une manière ou d'une autre, à l'instance y ? » en cherchant les chemins reliant les individus dans un graphe. Définie à l'origine pour des graphes RDF(S), nous avons appliqué l'analyse sémantique aux ontologies OWL (plus expressives) puis étendu ce processus d'analyse pour exploiter des informations spatiales et temporelles dans la déduction de nouvelles associations sémantiques mettant par exemple en exergue une proximité spatiale ou/et temporelle entre objets d'étude (des

³ Synchronized Multimedia Integration Language, désormais disponible en version 3.0, W3C Recommendation 01 December 2008 <https://www.w3.org/TR/2008/REC-SMIL3-20081201/>

⁴ Les références de type [TH-n] correspondent aux thèses pour lesquelles j'ai officiellement participé à l'encadrement. Elles sont listées en fin de manuscrit, page 116, avec mention des directeurs de thèses et financements.

⁵ W3C Web Ontology Language, <https://www.w3.org/OWL/>

personnes, par exemple). Ces travaux ont donc marqué le début de recherches menées dans le cadre du Web Sémantique. Ils se sont ensuite poursuivis, par exemple, dans le cadre du master de Qiang Fu, par l'évaluation de l'apport des techniques d'analyse sémantique et d'expression de contraintes en OWL pour la détection d'incohérences spatio-temporelles. L'objectif était de contribuer à une meilleure qualité de la donnée disponible, en identifiant les éventuelles inconsistances entre deux sources d'informations.

Approches orientées ontologies pour la représentation de connaissances

La question de la qualité des données du Web Sémantique a également été abordée dans la thèse de Anthony Hombiat [TH-3] (soutenue en février 2017, (Hombiat, 2017)), dans le contexte particulier dessiné par le nouveau mode de production des données qui fait de chaque internaute du Web un potentiel fournisseur d'informations. L'action de ces nouveaux capteurs (humains) de l'information (*citizen sensors* (Goodchild, 2007)) et de l'Information Géographique notamment, génère un flot massif de données sur des plates-formes de cartographie participative telles qu'OpenStreetMap (OSM⁶). Ayant largement impulsé le phénomène de l'Information Géographique Volontaire (*Volunteered Geographic Information*, VGI), la communauté OSM, forte de plus de deux millions de contributeurs, alimente une base de données géospatiales ouverte visant à capturer une représentation du territoire à l'échelle mondiale. Les *tags* associés aux objets géométriques du modèle OSM permettent une catégorisation simple et rapide des contenus, facilitant et donc favorisant l'action des contributeurs. Les *folksonomies* ainsi créées de manière massive souffrent en contrepartie d'un certain manque d'expressivité qui constitue un obstacle au partage et à la réutilisation de ces grands volumes d'information. Dans son travail de thèse, A. Hombiat explore comment structurer la *folksonomie* OSM pour augmenter le pouvoir sémantique sans nuire à sa flexibilité. Cela permet notamment de garantir une information actualisée, les contributions étant favorisées par la simplicité avec laquelle on peut taguer les contenus. L'approche proposée repose sur une ontologie appelée OF4OSM (*Ontology from Folksonomy for OSM*) qui est alimentée par un algorithme d'extraction et de traitement automatisés⁷ des tags qui caractérisent les objets géographiques de la base de données OSM. Le résultat est que les contributeurs n'ont pas à changer leurs habitudes de *tagging* mais que la sémantique de leurs tags, dont la richesse n'est pas explicite dans la *folksonomie* OSM, est révélée et augmentée dans l'ontologie OF4OSM. Cela est notamment possible car l'ontologie OF4OSM⁸ permet : 1) de construire automatiquement une hiérarchie de concepts à deux niveaux à partir des tags OSM originaux en exploitant leur structuration en paires clé-valeur, et 2) d'ajouter de nouveaux niveaux à la hiérarchie de concepts : a) en exploitant la factorisation de caractéristiques partagées par différents tags et b) en utilisant la conjonction de plusieurs tags associés à une même instance. L'approche permet ainsi d'améliorer la qualité des données OSM, au moment de l'acquisition, mais aussi lors de leur exploitation car l'ontologie permet de suggérer automatiquement des tags plus précis que ceux saisis par le contributeur, ou encore complémentaires à ces derniers (voir (Hombiat, 2017), Chapitre 7).

Les travaux que je présente dans ce document s'inscrivent dans la continuité de ces recherches en cela qu'ils visent à proposer des approches à base d'ontologies pour la représentation des connaissances. Une sélection a été faite à cet effet : un point commun des travaux décrits est qu'ils abordent la question de différentes formes d'évolution qu'il s'agit de représenter et d'exploiter, dans un contexte expert.

*« Le commencement de toutes les sciences, c'est l'étonnement de ce que les choses sont ce qu'elles sont » (Aristote).
Si les choses sont ce qu'elles sont, c'est aussi d'avoir été ce qu'elles ont été. Représenter l'évolution prend alors tout son sens.*

⁶ <https://www.openstreetmap.org/>

⁷ voir (Hombiat, 2017) pour l'implémentation

⁸ L'ontologie OF4OSM repose sur un méta-modèle également décrit dans (Hombiat, 2017)

“If you have an apple and I have an apple and we exchange these apples then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas.”

George Bernard Shaw (1856 – 1950)

Introduction

1

Cher lecteur, permettez-moi de donner à cette introduction une touche personnelle et de partager avec vous, d’abord, un état d’esprit. Longtemps je n’ai su comment envisager la rédaction de ce manuscrit compte-tenu d’une certaine hétérogénéité dans les travaux auxquels j’ai pris part. Depuis ma thèse différents domaines ont été explorés : les systèmes d’information, l’adaptation, le multimédia, la représentation de connaissances, le web sémantique, la géovisualisation, la cognition, et d’autres champs encore. Comment, dès lors, valoriser la variété des problématiques abordées sans donner l’impression d’une forme d’égarément ? Comment orchestrer ces travaux le long d’un fil directeur qui ne soit pas artificiel ? Où trouver dans tout cela la cohérence d’un parcours en recherche ? C’est à force d’échanges d’idées et de partages de points de vue qu’il m’est apparu que la valeur d’un travail de recherche peut aussi être dans la diversité des territoires explorés au cours du temps ; que la richesse d’un parcours est dans les trajectoires des vies croisés à l’occasion de ces différents travaux ; que des capacités se forment par la juste analyse des relations de cause à effet entre les événements vécus ; que les compétences s’acquièrent dans les interactions et les accompagnements obligeant à se réinventer pour être autant que possible en adéquation avec les différents interlocuteurs. C’est grâce à cela que la rédaction de ce document est devenue envisageable. Nourrie d’expériences, des plus anciennes aux plus récentes, il m’est devenu possible de prendre le recul nécessaire pour regarder mon parcours de recherche avec, non pas un esprit trop négatif, quelque peu paralysant, mais bien critique. Ironiquement, c’est la posture que je demande à mes doctorants d’adopter... Cette approche critique m’a amenée à l’acceptation d’un renoncement : ne pas chercher à tout prix à parler de tout. Alors, comme évoqué en préambule, je fais donc le choix dans ce document de présenter de façon plus détaillée certains seulement des travaux de recherche auxquels j’ai contribué. Pour autant, que tous les

chercheurs (en devenir, juniors ou seniors) avec lesquelles j'ai aussi collaboré dans le cadre de recherches non évoquées ici ne s'y trompent pas : il y a forcément d'eux dans ces écrits.

Il y a également dans ces écrits, en filigrane, un contexte dans lequel s'inscrivent mes recherches. Celui-ci est bien sûr marqué par les avancées scientifiques et techniques de la dernière décennie, parmi lesquelles les technologies sans fil, celles liées aux capteurs et celles ayant permis l'émergence du Web Sémantique. Il est aussi caractérisé par des changements sociétaux importants, avec un renouvellement des pratiques de l'informatique, du fait des réseaux sociaux ou encore des sciences participatives dont relève le *crowdsourcing*, par exemple. Ce contexte est, enfin, modelé par des initiatives, émanant de différentes instances de gouvernance, pour s'adapter à ce nouveau monde numérique et à ses utilisateurs (qu'ils soient humains – individuels ou institutionnels –, ou logiciels – agents, services, etc.) en promouvant des directives telles qu'Inspire⁹ au plan européen, ou encore en accompagnant la démarche Open Data¹⁰ jusqu'à un échelon local¹¹.

Le regard porté sur ce contexte mène à dresser plusieurs constats. D'une part, un déluge de données sans précédent est observé, et il va de pair avec un potentiel d'accessibilité à l'information accru. Les données, en partie libérées, sont atteignables de (presque) partout, (presque) par tout un chacun, pour (presque) tous les usages. Ces données renvoient à des questionnements qui alimentent les problématiques du *Big Data*, relatives au fait de collecter, stocker, explorer, partager, analyser ou encore visualiser ces *données aux 4V* (Volume, Variété, Vitesse et Vérité)¹². En particulier, pour le champ de la prise de décision, la capacité à « faire parler » ces masses de données est source de challenges à relever à de nombreux points de vue : technologique, économique, social, éthique, etc. La *Modern Artificial Intelligence* s'est ainsi nourrie de ce contexte, proposant des approches guidées par les données numériques (*numerical data-driven approaches*), parmi lesquelles, par exemple, l'apprentissage profond. Les approches symboliques, celles relevant de la *Good Old-Fashioned Artificial Intelligence* (GOFAI, d'après (Haugeland, 1985)), continuent néanmoins d'être exploitées et se réinventent pour s'adapter aux problématiques des données massives. Ces approches sont portées, notamment, par les logiques de description donnant à la représentation de connaissances à base d'ontologies et au Web Sémantique les fondements logiques nécessaires au raisonnement automatisé. De plus en plus, des travaux combinent *numerical data-driven approaches* et *symbolic approaches*, à l'image par exemple de (Al-Bakri et al., 2016) pour améliorer l'établissement de liens entre les données du web (*data linkage*), ou encore de (Banaee, 2018) pour construire des représentations conceptuelles à partir de jeux de données numériques issues de capteurs.

Pour autant, dans ce paysage, la compréhension de phénomènes (de nature sociodémographique, environnementale, sociologique, etc.) et la prise de décision peuvent, aussi, relever de l'exploitation d'informations qu'on ne peut pas qualifier de *Big Data* mais qui sont à l'origine de verrous scientifiques qui restent à lever. Parmi ces verrous, ceux qui m'intéressent particulièrement portent sur des questions relatives à la représentation (conceptuelle) de connaissances qui, sans toutefois être forcément *volumineuses* ni *véloces*, mais souvent *variées* et à la *vérité* questionnable, sont elles-aussi, versées dans le Web des Données Ouvertes et Liées ; par ailleurs, pour ces connaissances de nouvelles formes de représentation (géovisuelle) restent à inventer afin qu'elles expriment tout leur potentiel pour la compréhension de phénomènes et/ou la prise de décision.

Le cadrage scientifique de mes travaux est donc donné, d'une part, par la représentation des connaissances, immergées dans les ontologies du Web Sémantique et, d'autre part, par la géovisualisation. Ces deux champs sont explorés pour construire *les* représentations (conceptuelle et géovisuelle) de connaissances :

- caractérisées entre autres, par des dimensions spatiale et temporelle,
- potentiellement porteuses d'imperfection,
- dénotant une évolution des objets d'étude auxquels elles se rapportent et
- exploitées à des fins de compréhension de phénomènes et/ou de prise de décision.

⁹ <https://inspire.ec.europa.eu/>

¹⁰ <https://ec.europa.eu/digital-single-market/en/open-data>

¹¹ Voir par exemple les actions de l'association *OpenDataFrance*, <http://www.opendatafrance.net>.

¹² Les 4 V qui, selon IBM, pourraient bien être au nombre de 5, avec « *the ability to achieve greater Value through insights from superior analytics* », voir <https://www.ibmdatahub.com/infographic/extracting-business-value-4-vs-big-data>

Représentation de connaissances dans le Web des Données Ouvertes et Liées

Dans le Web Sémantique ou Web des Données Ouvertes et Liées (*Linked Open Data* ou *LOD*)¹³, les ontologies (Berners-Lee et al., 2001) exprimées en OWL et les graphes RDF¹⁴ notamment permettent la description de connaissances sous une forme compréhensible à la fois par des humains et par des machines. Elles offrent en cela à la fois un cadre conceptuel pour l'expression de faits et de règles, selon un vocabulaire partagé (ici, au sens de consensuel), et un cadre pour l'opérationnalisation informatique des faits et du raisonnement à partir de ceux-ci.

Du fait de la richesse de l'information disponible dans le Web des LOD, mais aussi de son accessibilité portée par les paradigmes et les technologies associées de l'*interopérabilité*, établir du lien entre connaissances décrites est facilité. Dans le champ de la représentation de connaissances, pouvoir partager (ici, au sens de mutualiser) des outils et des connaissances éprouvés est à la fois un gain de qualité et de temps. Les standards de description de caractéristiques spatiales comme GeoSPARQL¹⁵ ou temporelles comme OWL-Time¹⁶ sont des exemples d'outillage sur lesquels s'appuyer lorsqu'il s'agit de décrire et d'exploiter des connaissances de cette nature. L'utilisation de concepts ou d'instances décrits dans des ontologies de haut-niveau (*top-level ontologies*) ou fondamentales (*foundational ontologies*), ou des ontologies plus spécifiques dites « de domaines », « métiers » ou « de tâches », jusqu'aux ontologies « d'application » (Guarino and Welty, 2000) peut être envisagée pour décrire la connaissance qu'il est nécessaire de représenter dans une ontologie cible (dont les concepts et les instances pourront également être, à leur tour, ré-exploités).

Le Web des LOD est donc synonyme d'une ouverture sur des champs de connaissances connexes, complémentaires, ou apportant même des points de vue contradictoires, ce qui est une façon d'enrichir ses propres données, en les étendant, en les contextualisant, en les confrontant à des sources possiblement opposées. Dans un contexte qui vise à soutenir la compréhension de phénomènes (notamment en explicitant leur *évolution*), et/ou à faciliter à la prise de décision, à partir de connaissances décrites dans des ontologies, ce potentiel d'accès à d'autres données est à exploiter. Pour autant, allant dans le sens des propos de (Janowicz and Hitzler, 2013), nous pensons qu'il est important que les ontologies conçues et développées dans un but de compréhension de phénomènes et/ou de prise de décision le soient à un niveau conceptuel adapté à ces objectifs. Les auteurs de (Janowicz and Hitzler, 2013), évoquent l'idée suivante : « *Maybe, at least for Linked Data, it is time to give up on the idea of context-free ontologies as models of the physical world and instead define a multitude of purpose and data-driven micro-ontologies;* ». Ici, l'expression *data-driven* n'est pas à comprendre dans le sens des approches numériques guidées par les données comme dans la *Modern AI*, mais bien comme relevant des approches symboliques, avec une représentation conceptuelle des données qui guide le processus de construction de l'ontologie. Les auteurs poursuivent en effet en indiquant : « *Ontologies should be engineered based on the real data they are supposed to reflect and their axiomatization should be driven by the inference needs of typical queries.* ». Cela évoque l'importance, clairement identifiée dans l'univers des systèmes d'information et de l'ingénierie des besoins, de considérer l'utilisateur – humain ou non – et les usages.

La construction d'ontologies ciblées ne signifie pas qu'il faille abandonner tout objectif de *généricité* dans la description des connaissances. Si l'ontologie à créer sert certains objectifs en particulier, elle doit pouvoir aussi venir enrichir le Web des LOD (en tant que modèle ou par la mise à disposition d'instances utiles à d'autres cas d'application). Ainsi, trouver le bon niveau de conceptualisation doit être gardé à l'esprit pour trouver un juste équilibre entre la réponse à un contexte applicatif particulier et la contribution à la communauté. Nous cherchons à maintenir cet équilibre dans nos travaux.

¹³ Dans ce manuscrit, nous employons comme des synonymes les expressions *Web des Données Ouvertes et Liées*, *Web des LOD* ou simplement *LOD* pour faire référence au cinquième niveau du schéma de développement des *Linked Data* selon Tim Bernes-Lee <https://www.w3.org/DesignIssues/LinkedData.html>.

¹⁴ Resource Description Framework, <http://www.w3.org/RDF>

¹⁵ <http://www.opengeospatial.org/standards/geosparql>

¹⁶ <https://www.w3.org/TR/owl-time/>

Décrire un phénomène pour mieux l'appréhender et étayer une prise de décision nécessite de considérer les caractéristiques factuelles intrinsèques de ce phénomène, mais aussi d'intégrer dans la représentation les éléments susceptibles d'expliquer les faits observés. A cet égard, les ontologies dédiées à la représentation des phénomènes passés dont on souhaite comprendre l'évolution doivent être dotées d'une dimension *descriptive* et d'une dimension *explicative*. Il s'agit d'un second objectif que nous visons.

S'agissant de retranscrire l'évolution de phénomènes passés, de possibles *imperfections* existent dans les sources d'information disponibles. Retranscrire ces imperfections est crucial en vue de comprendre et de prendre des décisions. Considérer des solutions pour gérer différentes formes d'imperfection dans la représentation des connaissances à base d'ontologies constitue un troisième objectif.

Si les ontologies sont connues pour être des formalismes lisibles par l'humain, offrir une interface d'accès à la connaissance qui dépasse sa mise à disposition au format textuel est tout de même fort utile pour une meilleure appréhension et une exploitation plus riche de cette connaissance. Il s'agit d'un autre axe de recherche auquel je m'intéresse, à travers, plus particulièrement, la géovisualisation.

Géovisualisation

Revenons quelques instants sur les *Big Data* qui constituent cette matière première si présente aujourd'hui, en considérant qu'il ne s'agit plus de représenter conceptuellement cette vaste source d'informations mais de la *restituer* à un utilisateur. Les champs de la *Data Visualisation*, de la *Scientific Visualisation (SciVis)* (McCormick et al., 1987), ou de l'*Information Visualization (InfoVis)* (Card, 2008; Card et al., 1999) ont en commun de concevoir et développer les outils permettant de travailler avec de gros volumes de données. Si les approches diffèrent par les caractéristiques des données sources et les objectifs des restitutions (Chen, 2005), dans les environnements produits, il est généralement question de « donner à voir ou à explorer » des données et de favoriser « la découverte de l'inconnu » par la combinaison de différents modes de visualisation (Thomas and IEEE Computer Society, 2005), et, grâce aux interactions possibles de l'utilisateur avec l'outil. Lorsque les données présentent un caractère spatial et/ou temporel, on s'ouvre sur le champ de l'analyse géovisuelle (*Geovisual Analytics*, (Andrienko et al., 2007; Robinson, 2017)) et des *Geospatial Big Data* pour lesquels différents challenges sont identifiés pour la communauté (voir notamment (Robinson et al., 2017)). À l'image de ce que nous avons indiqué pour nos travaux en représentation des connaissances, nos recherches portant sur la restitution graphique de celles-ci ne sont pas caractérisées par la gestion du caractère massif et volubile des données, mais la littérature dans ces différents champs de recherche autour de la visualisation reste une source d'inspiration importante. Nos travaux se revendiquent plus clairement de la *géovisualisation (MacEachren and Kraak, 1997)* et en particulier dans qu'elle peut apporter à la construction de la connaissance et à la prise de décision (MacEachren et al., 2004) par la construction d'environnements dédiés. Dans (Christophe, 2017), l'auteur distingue la géovisualisation en tant qu'outil¹⁷ de la géovisualisation en tant que champ interdisciplinaire de recherche¹⁸. Nous prolongeons ce point de vue en ajoutant qu'un lien implicite existe entre les deux : la *géovisualisation-recherche* vise à proposer *in fine* des *géovisualisations-outils* utiles, utilisables, adaptés, efficaces, performants, etc. Pour (Davoine, 2014), « *tout l'intérêt de ces applications réside dans leurs capacités à donner de multiples éclairages, qu'il n'est pas possible de réunir au sein d'une seule représentation*

¹⁷ « La géovisualisation est un outil dynamique et interactif pour afficher des données géographiques, des modèles de données, des cartes, des images, et des données externes géo-localisables, permettant à un utilisateur de raisonner spatialement et/ou temporellement sur ces représentations, et d'inférer des connaissances sur le territoire ou le phénomène représenté, en offrant la capacité d'interagir avec une ou plusieurs dimensions du phénomène. » (Citation S. Christophe, 2017, p5)

¹⁸ « La géovisualisation est un champ interdisciplinaire de recherche qui s'intéresse aux techniques de visualisation d'un territoire ou d'un phénomène sur un territoire, avec des problématiques d'intégration visuelle de données géographiques hétérogènes (multi-sources, multi-échelles spatiales et temporelles) de représentation graphique, de rendu et d'interaction avec l'utilisateur, ainsi qu'à l'utilisabilité de ces techniques pour des usages potentiellement variés. » (Citation S. Christophe, 2017, p5)

cartographique ». Ces éclairages, souvent temporels et thématiques, sont retranscrits dans des composants d'interface synchronisés avec le composant, généralement central dans une géovisualisation, qui accueille la carte. Ce statut de la carte légitime d'ailleurs le préfixe « *géo* », mais de nombreux exemples de réalisation montrent que les autres dimensions de l'information, dont la dimension temporelle, sont aussi très présentes dans les *géovisualisations-outils* et des sujets d'études de la *géovisualisation-recherche* (voir des revues par exemple dans (Blaise et al., 2015; Davoine, 2014; Gautier, 2018; Saint-Marc, 2017; Zanin et al., 2013)).

Parmi les sujets de recherche en géovisualisation auxquels je contribue, certains font l'objet de questionnements partagés au sein de l'Action Prospective *Géovisualisation et Cognition*¹⁹ du Groupement De Recherche du CNRS MAGIS, que je co-anime avec S. Christophe (Christophe, 2017). Les principales orientations des réflexions menées ici relèvent de l'apport des sciences cognitives, et en particulier de la psychologie cognitive, dans le *processus de construction* des géovisualisations. Il s'agit, d'une part, d'intégrer ce que nous enseignent ces sciences à propos des capacités perceptives humaines pour proposer des formes de transcriptions visuelles des connaissances, et des modalités d'interactions associées, innovantes mais surtout efficaces. Un second axe de recherche porte sur l'aide qu'on peut apporter au concepteur, en outillant sa démarche d'un point de vue méthodologique pour faciliter la production de géovisualisations pertinentes, idéalement en favorisant de bonnes pratiques (dont l'identification reste également un challenge important dans la communauté).

Enfin, face à la multitude de visualisations possibles pour l'information géographique, une de nos préoccupations est de formaliser la connaissance et les stratégies qui sous-tendent le raisonnement humain pour mieux répondre aux besoins de l'utilisateur en situation de prise de décision, comme évoqué dans (Fabrikant, 2001). C'est autour de cette question que s'achèvera ce document.

Organisation du Document

Les représentations de connaissances spatiales et temporelles relatives de la notion d'*évolution* constituent le fil directeur de ce document. Trois approches sont présentées. Elles s'inscrivent chacune dans un contexte pluridisciplinaire particulier, et l'évolution dont il est question s'exprime de façon différente, du fait de la spécificité des objets d'étude pour lesquels la notion même d'évolution, décrite et expliquée, prend des formes différentes. Un chapitre est dédié à chacune de ces recherches.

Ainsi, dans le chapitre 2 nous étudions *l'évolution des territoires*, en nous appuyant essentiellement sur la thèse de Camille Bernard [TH-6], qui prolonge les travaux conduits par Christine Plumejeaud lors de son doctorat (Plumejeaud, 2011), réalisé au sein de notre équipe et auxquels j'ai collaboré (Plumejeaud *et al.*, 2009)(Plumejeaud *et al.*, 2011b). Notre objet d'étude est ici un territoire observé selon un découpage (par exemple, un découpage administratif en régions, départements, communes, *etc.*). L'évolution du territoire est plus exactement à comprendre comme celle de ce découpage qui peut subir des modifications au cours du temps, pour des raisons variées. L'approche ontologique que nous proposons permet de décrire les différentes versions de ce découpage et de les expliquer en donnant une représentation détaillée des changements entre les versions et en contextualisant ces changements grâce aux liens qu'il est naturellement possible d'établir avec des ressources du Web des Données.

Le chapitre 3 est centré sur un deuxième objet d'étude : l'individu, dont on suivra l'évolution dans le temps et dans différents espaces, à travers la représentation de sa *trajectoire de vie*. Ce travail fait principalement l'objet de la thèse de David Noël [TH-5]. Nous proposons ici une approche générique à base de patrons d'ontologies permettant la représentation multidimensionnelle (*i.e.* selon différents points de vue, en fonction des besoins) d'une trajectoire de vie qui intègre des facteurs explicatifs associés aux étapes et aux événements qui jalonnent cette trajectoire de vie.

Enfin, le travail réalisé par Cécile Saint-Marc dans le cadre de sa thèse [TH-4] a inspiré le chapitre 4, qui aborde *l'évolution d'une situation*, dans le domaine des risques, de la survenue de l'événement déclencheur de la crise, jusqu'au retour à la normal. Menée conjointement avec la SNCF dans le cadre d'une CIFRE, l'approche proposée est une contribution aux démarches de Retour d'Expérience. La description de la situation en elle-même s'accompagne de l'explicitation des liens de

¹⁹ <http://gdr-magis.imag.fr/actions-prospectives/geo-visualisation-cognition/>

causalité (de type effet dominos) entre les événements et/ou les décisions prises, favorisant ainsi la compréhension de ce qui s'est joué suite à l'événement initial. Ce travail est également celui à travers lequel nous exposons quelques unes de nos recherches en géovisualisation.

Chaque chapitre débute par une présentation du contexte scientifique dans lequel s'inscrit le travail, en présentant les motivations et les enjeux, suivis d'un rapide état de l'art. Nous organisons ensuite chaque chapitre en différentes sections qui visent à rendre compte des contributions principales. Les chapitres se concluent par une synthèse et une mise en perspective des travaux.

« [...] if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas. » (George Bernard Shaw). Le 5^{ème} et dernier chapitre de ce manuscrit dresse un bilan des activités de recherches qui sont venues donner corps à bien des idées échangées. Il est aussi l'occasion pour moi de partager avec le lecteur, à défaut d'échanger, quelques idées sur les orientations de mes travaux à venir.

« *None of us see the world as it is but as we are,
as our frames of reference, or maps, define the territory.* »

Stephen R. Covey (1932 – 2012)

Découpages territoriaux évolutifs 2

2.1 Contexte des travaux

2.1.1 Motivations et enjeux

Dès lors que l'on considère le territoire comme objet d'étude, la nécessité d'en évoquer les limites pour le définir vient à l'esprit, et ce quelle que soit la discipline (la géographie - sous toutes ses formes, l'urbanisme, la biologie, le marketing, la politique, *etc.*). Les limites d'un territoire sont à la fois celles de ses frontières, qui le distinguent d'autres territoires et celles, internes, qui le structurent. Un territoire peut ainsi être « découpé » selon des critères servant un objectif particulier, le plus souvent organisationnel ou politique quand il est dicté par l'Homme. Ainsi, une *Nomenclature Territoriale Statistique* (TSN pour *Territorial Statistical Nomenclature*) est un découpage du territoire, utilisé en statistique à des fins de collecte de données socio-économiques. De telles nomenclatures (comme, par exemple, les *Functional Urban Areas* (FUA)²⁰ ou *Nomenclature of Territorial Units for Statistics* (NUTS)²¹) varient par leur forme (*e.g.*, *zonage* couvrant ou non) ou par les espaces qu'elles décrivent (métropoles, pays, *etc.*), mais ont en commun de diviser les territoires en unités territoriales, parfois sur plusieurs niveaux (*e.g.*, pays – *niveau 0*, grandes régions – *niveau 1*, régions – *niveau 2* dans la NUTS). Dans ce travail, nous nous concentrons sur les TSN découpant le territoire en polygones irréguliers définissant une partition du territoire, possiblement à différents niveaux. La NUTS, définie à l'échelle de l'Europe est un exemple d'une telle nomenclature.

Ces découpages territoriaux peuvent varier au cours du temps : le nom d'une communauté de communes peut changer, ou bien les régions administratives d'un pays peuvent être redéfinies, comme en France en 2016. Un objectif de capitalisation des connaissances en soi est une motivation que l'on peut déjà juger suffisante pour garder la trace de ces évolutions. Le recours aux technologies du Web Sémantique, la richesse et l'accessibilité des informations de type *LOD*, permettent, en outre, de

²⁰ http://ec.europa.eu/eurostat/statistics-explained/index.php?title=Territorial_typologies_for_European_cities_and_metropolitan_regions

²¹ <http://ec.europa.eu/eurostat/web/nuts/overview>

contextualiser un changement territorial, par exemple en pointant vers une ressource décrivant la réforme. De même, en permettant la publication dans le Web Sémantique de la description des territoires dans leurs différentes versions, il devient possible de référencer le « bon » territoire pour ses données, celui qui est réellement le référentiel d'un indicateur donné (et pas uniquement le territoire selon la nomenclature actuellement en vigueur, qui peut ne plus être en adéquation avec la réalité au moment de la mesure ou du relevé de valeur de l'indicateur).

Au delà, en matière d'analyse spatiale par exemple, l'absence de traçabilité de l'évolution des découpages rend difficile la comparaison de données sur le long terme, puisque la constitution même d'une série temporelle (de statistiques socio-économiques par exemple) pour une unité territoriale est difficile à réaliser. En effet, les différents indicateurs (ou variables) comme la population, la richesse, *etc.*, sont issus du comptage des effectifs inclus spatialement dans cette unité. Puisqu'elle peut changer (de forme, de taille, de nom, *etc.*) au cours du temps, les valeurs de chaque variable ne sont pas toujours immédiatement et aisément comparables entre deux versions de cette unité, attachées à des versions successives de la nomenclature.

Les changements subis par les découpages territoriaux au cours du temps sont donc des informations capitales pour comprendre et travailler avec les dynamiques des territoires. En œuvrant pour la description et la traçabilité des changements dans les TSN, il est possible de répondre à différents objectifs opérationnels d'experts parmi lesquels :

a) donner à voir les changements de frontières au cours du temps, en prendre conscience et ainsi mieux comprendre les dynamiques des territoires par l'évolution de leurs limites internes ou de celles de leurs voisinages. Cet objectif fait référence à une chaîne de traitements cognitifs <visualiser – comprendre – raisonner – décider> sur laquelle nous revenons plus tard dans le manuscrit lorsque nous abordons la géovisualisation, domaine pour lequel disposer de représentations sémantiquement riches des connaissances est idéal. L'objectif est d'aller au delà de la visualisation de deux couches superposées d'informations correspondants aux délimitations des unités attachées à deux versions de nomenclature, comme proposé par exemple par l'outil ABS Map²².

b) produire automatiquement les valeurs d'indicateurs, socio-économiques ou environnementaux par exemple, dans une nouvelle version de la partition territoriale, en utilisant un programme capable de déterminer l'opération à appliquer aux données (*e.g.* agrégation, désagrégation ou autre estimation) en fonction de la nature de l'indicateur (*e.g.*, indicateur de type stock ou ratio) mais aussi de la nature du changement territorial en s'inspirant, par exemple, de travaux tels que ceux de (Flowerdew, 1991; Goodchild *et al.*, 1980).

c) simuler et observer les effets d'un redécoupage territorial à des fins de gouvernance, en lien par exemple avec la gestion des subventions. La fusion de deux communes peut avoir des répercussions sur la valeur du revenu moyen par habitant et impacter, à la baisse, le montant des subventions nationales reçues par une des communes mais, inversement, elle peut réduire considérablement le coût de collecte et de traitement des déchets.

d) disposer de trajectoires territoriales sur le temps long, afin d'étudier la pertinence de suivre ou non un chemin similaire en termes de recomposition, extension territoriale, *etc.* Les descriptions du changement peuvent permettre d'identifier des *patterns* d'évolution des territoires, et par suite d'identifier des territoires dont les trajectoires sont similaires ou opposées pour conduire des comparaisons ou des analyses d'impact des politiques menées.

2.1.2 Ontologies pour la structuration d'un territoire et ses évolutions

Une TSN, en tant qu'ensemble structuré de connaissances décrivant un territoire, peut se représenter sous la forme d'une ontologie. Le territoire décrit pouvant évoluer au cours du temps, une TSN existe généralement en plusieurs versions, datées, chacune visant à rapporter des changements territoriaux survenus au cours du temps. Dans le contexte des ontologies, les *changements du domaine* ou *du monde réel* doivent être distingués des *changements de conceptualisation* (Noy and Klein, 2004). Les modifications subies par le découpage territorial relèvent du premier type (Kauppinen and Hyvönen, 2007) : ce n'est pas la TSN en tant que le modèle de description qui change avec une nouvelle version, mais bien les instances de ce modèle. Les concepts de description de l'ontologie,

²² Outil proposé par l'*Australian Bureau of Statistics*, <http://stat.abs.gov.au/itt/r.jsp?ABSMAPS>

leurs propriétés et relations restent les mêmes d'une version à l'autre, alors que les instances de ces concepts, les valeurs de propriétés et de relations peuvent, elles, évoluer.

Bien qu'à travers le monde, les territoires sont généralement structurés (administrativement ou encore à des fins de collecte statistique) de façon assez similaire dans le principe (*i.e.* en unités territoriales constituant une partition et agrégables en unités de niveaux supérieurs), il n'existe pas une seule TSN générique mais plusieurs, issues de différentes initiatives. En Europe, la NUTS déjà évoquée, offre un découpage de l'ensemble du territoire, en 4 niveaux²³. Elle est révisée environ tous les 3 à 4 ans, chaque version rendant compte des modifications survenues au cours de la période écoulée depuis la version précédente. De façon similaire, le territoire australien a été décrit de 1984 à 2011 par une nomenclature (*Australian Standard Geographical Classification*, ASGC)²⁴ révisée annuellement et disponible au format *csv*. En 2011, cette nomenclature est abandonnée au profit d'une nouvelle appelée *Australian Statistical Geography Standard* (ASGS)²⁵, dont une deuxième version est proposée en 2016. D'autres instituts statistiques adoptent une approche ontologique pour la description et la publication de leur TSN dans le Web des LOD :

– l'INSEE a développé une *Ontologie géographique*²⁶ utilisée pour publier au format RDF les données du Code Officiel Géographique²⁷. Les concepts sont décrits en français et sont extrêmement dépendants du contexte territorial national (*e.g.*, concept *ArrondissementMunicipal*). Dans une version mise en ligne fin 2017 / début 2018, l'ontologie proposée intègre un support accru pour décrire le changement, sans cependant couvrir l'ensemble des types d'évolution territoriale possibles, ni offrir les moyens d'une description fine de ces changements.

– L'*Ordnance Survey* (agence gouvernementale chargée de la cartographie du Royaume-Uni (UK)) publie l'ontologie *Administrative geography and civil voting area ontology for UK*²⁸. Elle permet de décrire les niveaux de la NUTS, mais la dimension temporelle des objets géographiques est oubliée (Correndo *et al.*, 2010). Là aussi, les concepts définis dépendent fortement du contexte britannique.

– L'Office National pour la Statistique (ONS) du UK, propose plusieurs vocabulaires²⁹ pour représenter un espace géographique et son évolution au cours du temps (Ontologie *ONS Boundary Change Ontology*³⁰). Un certain niveau d'abstraction dans le choix des termes est à noter, notamment pour la description des unités territoriales³¹. Toutefois, aucune de ces solutions ne permet de décrire la structure des TSN et les niveaux territoriaux qui les composent. Dès lors, si l'on souhaite décrire d'autres découpages territoriaux que ceux listés dans l'ontologie *Geographical hierarchy ontology*³² (*e.g.*, *European Electoral Region, Parish, etc.*), il est nécessaire d'ajouter de nouveaux concepts à cette ontologie (*i.e.* de modifier le modèle). Concernant la description des changements dans les découpages au cours du temps, aucun des concepts de l'*ONS Boundary Change Ontology* ne permet de nommer des changements territoriaux qui impactent plusieurs unités territoriales en même temps.

²³ Désormais de NUTS 0 (les états) à NUTS 3 (les départements en France), les niveaux NUTS 4 et NUTS 5 n'étant plus intégrés à la nomenclature NUTS mais gérés dans le système des Unités Administratives Locales (LAU), voir <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>.

²⁴ [http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Standard+Geographical+Classification+\(ASGC\)](http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Standard+Geographical+Classification+(ASGC))

²⁵ [http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+\(ASGS\)](http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS)). L'ASGS vise une plus grande stabilité dans la description des territoires que ne le permettait l'ASGC et sépare désormais à cet effet les unités territoriales définies par l'ABS (*ABS Structures*, annoncées stables jusqu'en 2021, date du prochain recensement) et pour lesquelles il produit des informations statistiques, des unités territoriales non définies ou maintenues par le bureau (*Non-ABS Structures*).

²⁶ <http://rdf.insee.fr/def/geo/insee-geo-onto.ttl>

²⁷ <http://www.insee.fr/fr/information/2016807>

²⁸ <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

²⁹ <http://statistics.data.gov.uk/vocabularies>

³⁰ <http://statistics.data.gov.uk/def/boundary-change>

³¹ <http://statistics.data.gov.uk/def/statistical-entity>

³² <http://statistics.data.gov.uk/def/hierarchy/best-fit>

Plus généralement, très peu de concepts sont définis pour décrire la nature des changements territoriaux (*i.e.*, ils se limitent à *Recoding change*, *Boundary change*).

– Le Bureau des Statistiques du Japon propose des vocabulaires³³ pour représenter des régions géographiques mais aussi leurs évolutions au cours du temps. Des termes permettent de décrire les changements subis par ces régions, dans le contexte de la publication de données statistiques, reposant sur ces régions évolutives. Dans (Yamamoto *et al.*, 2017), les auteurs proposent une solution qui rejoint nos motivations pour la description dans le Web des LOD des changements subis par des régions géographiques et les raisons de ces changements. Néanmoins, ces descriptions se limitent aux changements dans les municipalités, en ayant recours aux concepts suivants : *absorption*, *abolishment*, *separation*, *establishment*, *division*, *name change*, *boundary change* et *shiftToAnotherKindOfCity*. D'autres étiquettes pour décrire les changements territoriaux sont cependant nécessaires. De même, une représentation multi-niveaux des changements, partant d'un changement père qui impacte des sous-éléments de la TSN, est à privilégier pour considérer ensemble tous les changements territoriaux qu'il fait sens d'analyser conjointement, à chaque sous-niveau du découpage.

– L'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) publie un jeu de données contenant une liste de pays du monde (décrits en RDF grâce aux concepts de l'ontologie *FAO Geopolitical Ontology*³⁴ tels que *self-governing*, *non-self-governing*, *disputed*) et les versions successives de ces pays (via les prédicats *predecessor*, *successor*, *valid since*, *valid until*). Cependant, les changements subis par un pays, d'une version à l'autre, ne sont pas décrits. Du fait du niveau de granularité retenu (celui des Etats), les changements dans les découpages internes des pays ne sont pas non plus décrits.

– Eurostat³⁵ propose l'ontologie *Reference And Management Of Nomenclatures Ontology (RAMON)*³⁶ qui permet la description des TSN NUTS et *Local Administrative Units (LAU)*³⁷. Cependant, pour décrire des unités territoriales provenant d'une autre TSN que la NUTS ou LAU, il est nécessaire de modifier l'ontologie et d'ajouter un sous-concept au concept *Geographical Region*.

Concernant plus particulièrement le support des différentes versions de la NUTS (que nous utilisons comme cas d'étude dans la suite), on peut noter que :

– Le jeu de données nommé *NUTS-RDF Geovocab*³⁸ est publié par le projet *Planet Data EU Network of Excellence*³⁹. Ici, seule une version de la NUTS (pas forcément la dernière) est publiée en ligne ne permettant pas le suivi des évolutions de la NUTS au cours du temps donc.

– *Eionet* publie au format RDF des versions de la NUTS⁴⁰ décrites via l'ontologie *RAMON* dont il propose une version simplifiée⁴¹. Des liens sont établis entre les unités territoriales des versions 2006 et 2003 (uniquement) de la NUTS : un lien *owl:sameAs* est utilisé entre des unités territoriales inchangées d'une version à l'autre, et un lien *rdfs:seeAlso* relie celles qui ont subi un changement. Les changements subis d'une version à l'autre ne sont, quant à eux, pas décrits.

– Le projet *EnAKTing*⁴² publie plusieurs versions de la NUTS au format RDF, sous le nom *Linked NUTS* (Correndo and Shadbolt, 2013). Les *Linked NUTS* décrivent les modifications subies par les unités territoriales de la NUTS, au fil des versions, mais la typologie des changements utilisée est restreinte (changement de code d'une UT, fusion, scission de UT).

³³ *e.g.* <http://data.e-stat.go.jp/od/sac/>, <http://data.e-stat.go.jp/od/terms/sacs#>, <http://data.e-stat.go.jp/od/sace/>

³⁴ <http://www.fao.org/countryprofiles/geoinfo/en/>

³⁵ Eurostat est l'office de statistique de l'Union européenne. <http://ec.europa.eu/eurostat/fr/home>

³⁶ <http://ec.europa.eu/eurostat/ramon/ontologies/geographic.rdf>

³⁷ <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>

³⁸ <http://nuts.geovocab.org/>

³⁹ <http://www.planet-data.eu>

⁴⁰ <http://rdfdata.eionet.europa.eu/ramon/nuts.rdf>; <http://rdfdata.eionet.europa.eu/ramon/nuts2008.rdf> ; <http://rdfdata.eionet.europa.eu/ramon/nuts2003.rdf>

⁴¹ <http://rdfdata.eionet.europa.eu/ramon/ontology.rdf>

⁴² <http://gtr.ukri.org/projects?ref=EP%2FG008493%2F1>

2.1.3 Positionnement de l'approche

Si l'on considère la capacité à décrire la structure définie par une TSN (en tant que découpage en unités territoriales en vigueur à un temps t), aucune des ontologies présentées ci-dessus n'a été proposée en cherchant à se placer à un niveau d'abstraction suffisant pour décrire n'importe quelle TSN. Les contextes nationaux sont souvent déterminants dans la façon de structurer les concepts dont les noms sont choisis pour être proches du « terrain » et dans la langue du pays. Disposer d'un modèle de description unifié (dans le choix des concepts, des relations et des vocabulaires) permet de faciliter la compréhension d'une nouvelle nomenclature puisqu'elle serait définie à partir d'éléments consensuels. Cela simplifierait la manipulation conjointe de différentes nomenclatures dont le référentiel serait de fait commun (réduisant ici la nécessité d'avoir recours à des techniques d'alignements entre ontologies). De plus, quelle que soit la TSN, l'exploitation de ses connaissances pourrait bénéficier de méthodes prédéfinies implémentées sur le modèle unifié, et, par suite, applicables à toutes ses instances. Une première contribution des travaux menés dans le cadre de la thèse de C. Bernard concerne donc la proposition d'une ontologie générique pour la description des TSN dans le Web des LOD (voir section 2.2) et (Bernard et al., 2018b, 2017a).

Le deuxième axe du travail concerne la description de l'évolution de ces TSN (voir section 2.3). Il est ici nécessaire d'explicitier les liens existants entre les unités territoriales de TSN définies aux temps t et $t+1$. L'enjeu est ici de choisir une représentation qui limite la redondance d'information mais permet une richesse sémantique importante, notamment pour ce qui relève de la description des changements survenus.

Pour représenter l'évolution des unités territoriales une approche de type *fluent* (voir Figure 1, approche (1)) peut être adoptée. Ici, les attributs porteurs de l'identité d'un objet O , s'ils varient, entraînent la création d'un nouvel objet. Ces attributs sont distingués des autres attributs qui, s'ils varient, entraînent la création d'un nouveau sous-objet de l'objet principal. Appelé *TimeSlice*, ce sous-objet, porteur de tous les attributs de l'objet variables dans le temps, a une période de validité définie (Welty et al., 2006). Cette approche est souvent utilisée dans le Web des LOD, en complément de l'ontologie *OWL-Time* pour représenter l'évolution des ressources (Batsakis, Petrakis, 2011), notamment géospatiales (Harbelot et al., 2015; Tran et al., 2015). Dans des approches telles que celle de (Kauppinen et al., 2008; Kauppinen and Hyvönen, 2007), l'état complet d'une unité territoriale (*i.e.* les valeurs de toutes ses propriétés et relations) est décrit avant et après un changement, ce que critiquent les auteurs de (Lacasta et al., 2014) pour la prolifération des instances à laquelle cela mène lorsque les unités évoluent fréquemment. Ils préfèrent alors une approche qui consiste à estampiller temporellement chaque propriété d'objet pour indiquer sa période de validité. La question de la prolifération des instances est formulée également concernant l'approche *fluent* (Batsakis, Petrakis, 2011), les *TimeSlice* étant de nouvelles ressources attachées à l'objet principal, créées dès lors qu'un seul des attributs non-identitaires change, dupliquant de fait les autres attributs non-identitaires inchangés. Néanmoins, s'agissant de nouvelles connaissances à représenter, quelle que soit l'approche, la création de ressources supplémentaires est inévitable.

Concernant la description des changements, l'ontologie *Change Ontology*⁴³ de (Kauppinen et al., 2008; Kauppinen and Hyvönen, 2007), introduit la notion de *Change Bridge* pour relier des unités territoriales entre elles et signifier la nature du changement entre deux versions du territoire (voir Figure 1, approche (2)). Cette réification de la notion de changement présente l'avantage de pouvoir associer des descripteurs sémantiques (propriétés, relations) qu'il est plus difficile d'exprimer par un lien direct entre unités territoriales.

Dans notre approche, nous représentons les unités existantes dans toutes leurs versions, et nous décrivons les changements survenus de version à version (*Materialization Using Versions and Deltas*, selon (Stefanidis, 2014)). Nous choisissons l'approche *fluent* pour la description des évolutions des objets, à ceci près que nous privilégions le vocabulaire des statisticiens et acteurs du territoire (voir, Figure 1 approche (3)). Ainsi, là où les ontologies *fluent* utilisent le terme *TimeSlice*, nous utilisons le terme *Version*. Par ailleurs, l'approche de (Kauppinen et al., 2008) est retenue pour établir des correspondances entre versions successives d'une TSN via des *Change Bridge*, mais nous enrichissons les descriptions des changements territoriaux à l'intérieur de ces *Change Bridge*, par l'expression des

⁴³ <http://linkedearth.org/change/ns/>

causes du changement. De plus, les travaux qui abordent la description des changements ne considèrent qu'une typologie réduite (c'est le cas de l'INSEE ou du projet EnAKTing) et lorsqu'elle est plus élaborée, la typologie des changements n'est définie que pour certains type d'unités (voir les propositions du Bureau des Statistiques du Japon). Dans (Plumejeaud, 2011), la *Typologie des changements territoriaux* proposée pour les TSN est plus riche et distingue des situations que d'autres travaux assimilent à un seul et même cas.

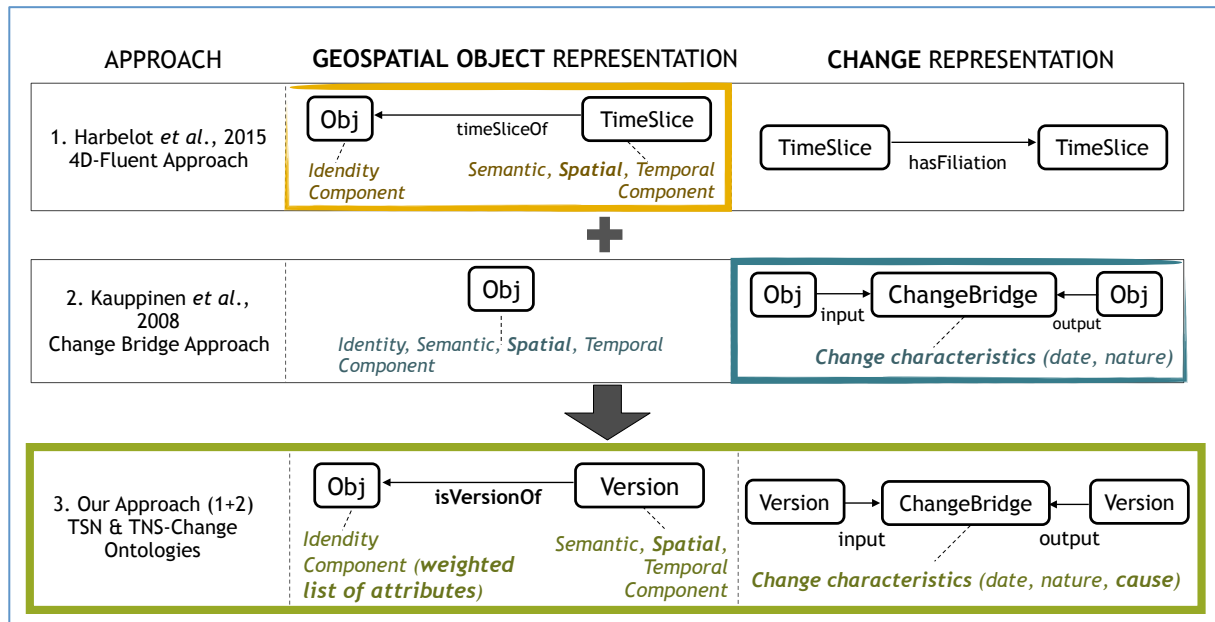


Figure 1 - Association de deux approches pour les ontologies TSN et TSN-Change

Auteur : Camille Bernard, 2018.

Dans ce qui suit, nous présentons les contributions pour la représentation de territoires évolutifs. Nous proposons ainsi, immergée dans le Web des LOD, une représentation détaillée et multi-niveaux des changements territoriaux ayant lieu au sein d'une TSN. Nous abordons dans un premier temps la question de la description de la structure d'un territoire, telle que connue au temps t et se conformant à une nomenclature qui dicte les modalités du découpage territorial (voir section 2.2). L'ontologie *TSN*, pour *Territorial Statistical Nomenclature*, a cette vocation descriptive. Dans un second temps, nous présentons les solutions offertes pour une représentation sémantiquement riche des changements observés entre deux versions d'un même territoire (voir section 2.3). L'ontologie *TSN-Change* est présentée à cet effet.

Techniquement parlant, les deux ontologies *TSN* et *TSN-Change* sont décrites via le langage d'ontologie *OWL* (*Web Ontology Language*), lui-même reposant sur une représentation des connaissances en triplets RDF $\langle \text{subject}, \text{predicate}, \text{object} \rangle$. Il s'agit d'ontologies géographiques qui définissent des concepts géographiques (e.g., une ville) et non géométriques (Klien and Probst, 2005). Conformément à la méthodologie de construction d'une ontologie de (Bachimont et al., 2002), le vocabulaire utilisé (pour être proches des désignations employées par les acteurs, et ainsi faciliter l'utilisation du modèle) a été déterminé à partir d'un corpus de TSN et ontologies existantes (basé essentiellement sur celles évoquées en 2.1.2). Nous montrons ensuite que l'automatisation de la découverte des changements entre deux versions de territoire est possible grâce à un algorithme inspiré des travaux de (Plumejeaud, 2011) et adapté à l'approche TSN/TSN-Change (voir section 2.4). Des exemples d'exploitation des connaissances produites sont donnés en section 2.5. L'ensemble de ces propositions est illustré par une application au cas de la NUTS. Ce chapitre se conclut par une synthèse et une ouverture sur des prolongements possibles de ce travail (voir section 2.6).

2.2 Représenter la structure du territoire

2.2.1 L'ontologie TSN

La Figure 2 présente, par des graphes RDF simplifiés, les principales classes et prédicats de l'ontologie TSN. Les *prédicats* apparaissent le long des flèches. Le départ d'une flèche indique le *domaine* et l'arrivée, le *co-domaine*. Les définitions des concepts sont accessibles à l'URI <http://purl.org/net/tsn>.

Les éléments de la Figure 2 reproduisent la hiérarchie des éléments composant une TSN selon une approche à base de versions. Les concepts centraux de *NomenclatureVersion*, *LevelVersion* et *UnitVersion* illustrent la structuration générique d'une nomenclature : les unités territoriales appartiennent à des niveaux qui permettent d'observer le découpage à des granularités différentes (*i.e.* selon des divisions plus ou moins fine de l'espace). Le prédicat *hasDivision(x, y)*⁴⁴ signifie "la *NomenclatureVersion* *x* est observable à un niveau de division *LevelVersion* *y*". Le prédicat *hasMember(x, y)* signifie "le *LevelVersion* *x* a comme membre la *UnitVersion* *y*".

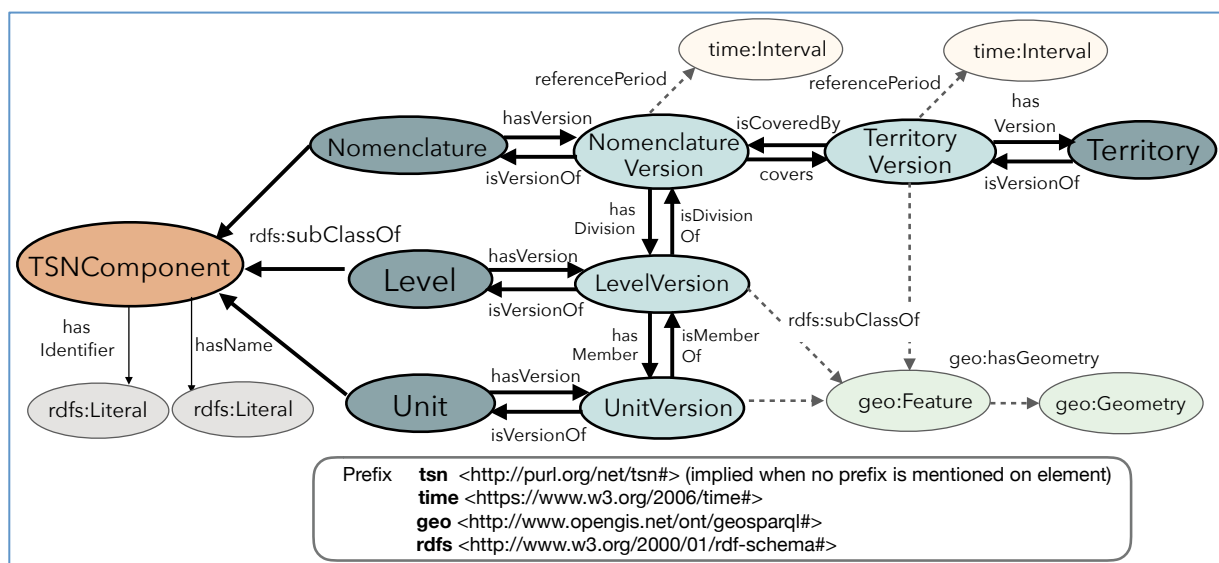


Figure 2 - Principales classes et prédicats associés de TSN Ontology

Source : <http://purl.org/net/tsn>, Auteur : Camille Bernard, 2018.

Ces concepts centraux sont à voir comme des instantiations concrètes de représentations plus générales dont on dit qu'elle ont des versions : *hasVersion(x, y)* signifie "le *TSNComponent* *x* a une *Version* *y*". Par exemple, le concept de *Nomenclature* peut exister en plusieurs *NomenclatureVersion* : ainsi, la NUTS (nomenclature abstraite) existe à ce jour en 6 versions (voir Figure 3). Une première version élaborée dans le cadre d'accords informels appelée NUTS 1999, suivie de 5 versions de statut juridique : NUTS 2003, NUTS 2006, NUTS 2010, NUTS 2013 et NUTS 2016. Chaque version est valide sur une période temps de 3 à 4 ans (l'année associée à la version correspond à l'année pour laquelle le découpage a été observé et collecté auprès des Etats Membres).

⁴⁴ Pour chaque prédicat mentionné dans cette section, une propriété inverse est définie.

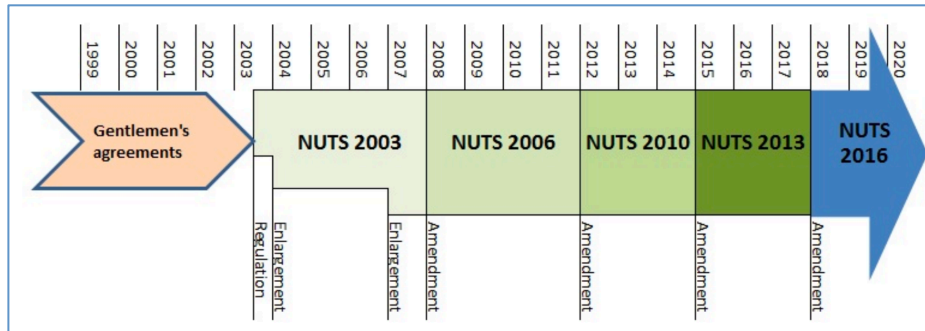


Figure 3 - Historique de la NUTS

Source : <http://ec.europa.eu/eurostat/web/nuts/history> (consulté le 20/09/2018)

Une version de nomenclature couvre une ou plusieurs versions de territoires (*TerritoryVersion*, des classes qui représentent les différentes versions d'un territoire *Territory*). Le prédicat *isCoveredBy(x, y)* signifie "la *TerritoryVersion* *x* est couverte par la *NomenclatureVersion* *y*". D'une part, cela permet de considérer qu'une *Nomenclature* (et par suite ses *NomenclatureVersion*) peut s'appliquer à un territoire dont les frontières externes évoluent. Ainsi, la NUTS découpe le territoire économique de l'Union Européenne (*Territory*) dont les limites au cours du temps se sont modifiées, le plus souvent du fait d'élargissements (des 6 pays de départ au 28 recensés à ce jour, ce qui a donné lieu à différentes *TerritoryVersion*), mais potentiellement aussi en lien avec des sorties du périmètre, comme voté par le Royaume-Uni en juin 2016. D'autre part, une même nomenclature peut couvrir plusieurs territoires comme c'est le cas pour la NUTS qui distingue le territoire des Etats Membres, de ceux des pays candidats (*Candidates Countries, CC*) et des pays membres de l'Association européenne de libre-échange (*European Free Trade Association, EFTA*). Enfin, la représentation que nous proposons ici permet également d'observer un même territoire selon différents découpages, comme par exemple en ayant recours au système d'unités administratives locales (LAU)⁴⁵ défini pour répondre à la demande de statistiques au niveau local.

Les concepts *TerritoryVersion*, *LevelVersion* et *UnitVersion* sont des sous-classes de *TSNFeature*, classe elle-même sous classe du concept *geo:Feature* défini par l'OGC dans l'ontologie *GeoSPARQL*⁴⁶. Le vocabulaire *OWL-Time* est utilisé pour assigner une période de validité au concept *NomenclatureVersion* et *TerritoryVersion*. Dans la hiérarchie des composants d'une version de nomenclature, cette période n'est assignée qu'à l'élément père (*NomenclatureVersion*) et s'applique par propagation aux éléments fils. Les prédicats (non illustrés dans la Figure 2) *hasLowerLevelVersion* et *hasUpperLevelVersion* expriment une relation hiérarchique entre *LevelVersion* d'une TSN. De même, *hasSubUnitVersion* et *hasSuperUnitVersion* traduisent les relations d'inclusion entre unités territoriales.

2.2.2 Cas d'étude : Description des composants de la NUTS

Les versions de la NUTS 2003 à 2016 sont disponibles sur le Web sous la forme de fichiers vectoriels géospatiaux⁴⁷, contenant la liste des unités pour chaque version ainsi que des attributs (identifiant dans la NUTS, nom, etc.). Pour transformer ces fichiers vecteurs en graphes RDF conforme à l'ontologie *TSN* nous avons⁴⁸ :

1. créé un fichier de *mapping* écrit en R2RML⁴⁹ permettant d'aligner les colonnes des fichiers vecteurs avec les concepts des ontologies *TSN* et *GeoSPARQL* (pour les géométries des unités).

⁴⁵ <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>

⁴⁶ <http://www.opengis.net/ont/geosparql>

⁴⁷ <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>. Au 16/07/2018, la version NUTS 1999 n'est plus téléchargeable, en revanche les versions 2013 et 2016 le sont désormais. Le travail présenté dans ce chapitre a été réalisé pour les versions 1999 à 2010 qui étaient disponibles au moment des réalisations.

⁴⁸ La procédure décrite peut s'appliquer à n'importe quelle TSN, dès lors que des fichiers vectoriels géospatiaux, contenant la liste des unités et leurs attributs (identifiant, nom, géométries and niveaux territoriaux dans la TSN) sont disponibles.

⁴⁹ <https://www.w3.org/TR/r2rml/>

- génééré automatiquement un graphe RDF pour chaque version de la NUTS grâce à l'outil *GeoTriples*⁵⁰ et son module *Mapping Processor*, en lui fournissant en entrée, le fichier vectoriel géospatial d'une version de NUTS et le fichier de *mapping* précédent.

La démarche a été appliquée pour les versions NUTS 1999, NUTS 2003, NUTS 2006 et NUTS 2010, ainsi que sur les données des versions 2011 et 2016 de la TSN australienne ASGS (cf. page 13) afin de tester la généricité de notre approche. Concernant les données de la NUTS, les graphes créés sont disponibles en ligne depuis le SPARQL *Endpoint* à l'adresse <http://purl.org/steamer/nuts>. Plus de 122 000 ressources ont ainsi été générées et publiées dans le Web des LOD dans la forme illustrée par l'Extrait de code 1 qui donne la représentation RDF de l'unité territoriale ES63 dans la version 1999 de la NUTS. Ces ressources proposées pour chaque version de la NUTS, et à chaque niveau de celle-ci, donnent la description de toutes les unités territoriales (voir lignes 7 à 16 pour un exemple de leurs propriétés, et lignes 19 à 30 pour la description de la géométrie).

Pour la publication de nos jeux de données NUTS dans le Web des LOD, les recommandations du W3C ont été suivies, et notamment celles abordant les questions de gestion des versions des données dans le Web – partie *Data versioning*⁵¹. Un identifiant unique (URI) associé à chaque élément de la NUTS (territory, level, unit, *etc.*) a été créé (*e.g.*, pour l'unité territoriale ES63 dans la version 1999 de la NUTS : http://purl.org/steamer/nuts/V1999_L2_ES63). Le service <https://purl.org/> a été utilisé pour créer des URIs persistants selon les patrons de construction suivants (voir Extrait de code 2) qui garantissent l'unicité des objets.

```
1@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2@prefix tsn: <http://purl.org/net/tsn#> .
3@prefix tsnchange: <http://purl.org/net/tsn#> .
4@prefix nuts: <http://purl.org/steamer/nuts/> .
5@prefix geo: <http://www.opengis.net/ont/geosparql#>
6
7nuts:V1999_L2_ES63
8  a          tsn:UnitVersion ;
9  tsn:hasIdentifier "ES63"^^xsd:string ;
10 tsn:hasName "Ceuta y Melilla" ;
11 tsn:hasSuperFeature nuts:V1999_L1_ES6 ;
12 tsn:isMemberOf nuts:V1999_L2 ;
13 tsn:isVersionOf nuts:L2_ES63 ;
14 tsnchange:hasNextVersion nuts:V2003_L2_ES63 ;
15 <http://dbpedia.org/ontology/languageCode> "es"^^xsd:string ;
16 geo:hasGeometry nuts:Geometry_3245 .
17
18nuts:Geometry_3512
19  a          geo:Geometry ;
20  geo:asWKT
21    "<http://www.opengis.net/def/crs/EPSG/0/4326>
22    MULTIPOLYGON (((-2.84 35.2937, -2.913 35.2633,
23    -2.9522 35.3492, -2.8959 35.3657, -2.893 35.3619,
24    -2.8593 35.3185, -2.84 35.2937)),((-5.2888 35.7974,
25    -5.3748 35.7767, -5.4141 35.8383, -5.334 35.8586,
26    -5.2888 35.7974)))"^^geo:wktLiteral ;
27  geo:is3D "false"^^xsd:boolean ;
28  geo:isEmpty "false"^^xsd:boolean ;
29  geo:isSimple "true"^^xsd:boolean ;
30  geo:spatialDimension 2 .
```

Extrait de code 1 - Description automatiquement générée d'une unité territoriale selon l'ontologie TSN

Cas de l'unité ES63 de la NUTS 1999, syntaxe RDF-Turtle⁵² – Auteur : Camille Bernard, 2018.

⁵⁰ <https://github.com/LinkedEOData/GeoTriples/wiki>

⁵¹ <https://www.w3.org/TR/dwbp/>

⁵² <https://www.w3.org/TR/turtle/>

```

1{base} TSN URI base (e.g., http://purl.org/steamer/nuts/ or @prefix nuts:)
2{base}V{ID TSN Version} TSN (e.g., nuts:V1999)
3{base}V{ID TSN Version}_L{ID Level} Level (e.g., nuts:V1999_L2)
4{base}V{ID TSN Version}_L{ID Level}_{ID UT} TU (e.g., nuts:V1999_L2_ES63)

```

Extrait de code 2 - Patrons de construction des URIs des éléments d'une TSN

Auteur : Camille Bernard, 2018.

Du fait de la notion de version qu'elle exploite (voir Figure 2, prédicat *isVersionOf* et un exemple de son utilisation dans l'Extrait de code 2, ligne 13), l'ontologie TSN permet de retrouver les différentes versions d'une unité territoriale, d'un niveau, d'une nomenclature ou encore d'un territoire au cours du temps. Enfin, nos travaux augmentent, au-delà de l'explicitation des relations de successions entre versions, les possibilités de description des changements qui permettent de caractériser l'évolution d'un territoire.

2.3 Représenter l'évolution d'un territoire

Représenter finement l'évolution d'un territoire requiert de pouvoir caractériser les changements survenus d'une version à l'autre de ce territoire, versions décrites via l'ontologie TSN. Nous proposons une approche originale qui se distingue de l'état de l'art par : *i*) le recours à une typologie des changements plus détaillée et versée dans le web des LOD, *ii*) la capacité à exprimer les liens entre changements impactant des niveaux différents de la nomenclature et *iii*) la possibilité de décrire des changements touchant un territoire d'étude dans ses frontières extérieures. L'ontologie TSN-Change, présentée ci-après, est le support de cette approche.

2.3.1 L'ontologie TSN-Change

2.3.1.1 Typologie des changements

La typologie des changements que nous proposons est illustrée par la Figure 4.

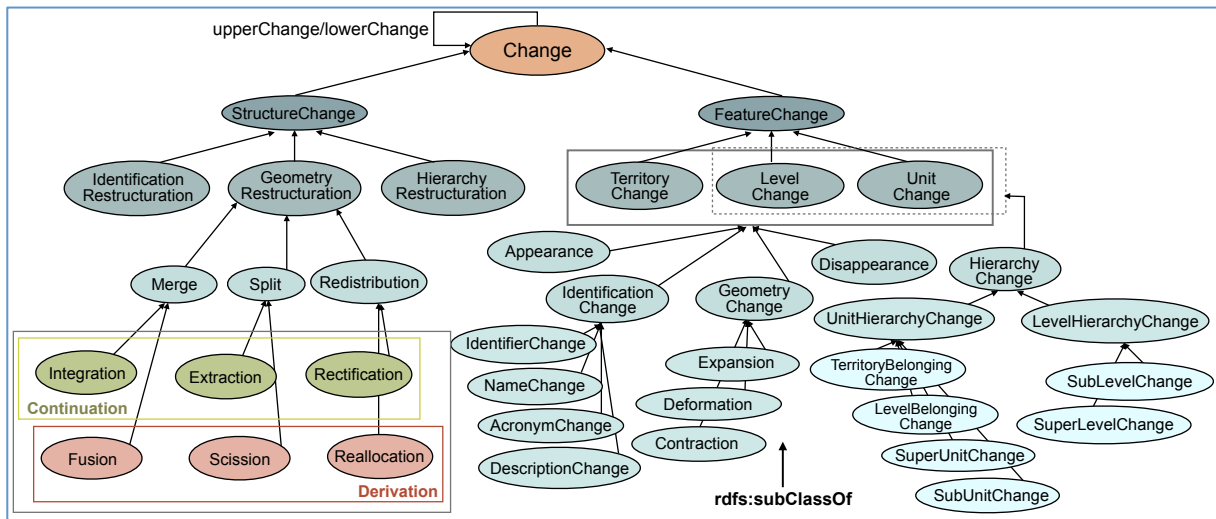


Figure 4 - Principales classes de l'ontologie TSN-Change

Source : <http://purl.org/net/tsnchange>, Auteur : Camille Bernard, 2018.

Deux grands types de changements sont distingués :

- *FeatureChange* est un concept utilisé pour décrire un ou plusieurs changements qui ont la particularité de ne toucher qu'un seul élément de type *TerritoryVersion*, *LevelVersion* ou *UnitVersion*. Il permet de mettre l'accent sur le suivi de l'évolution du point de vue d'une unité territoriale en particulier, par exemple. Ce concept se spécialise en différentes sous-classes permettant de décliner les différents types de changement.

- *StructureChange* est un concept dont la vocation est de décrire des restructurations qui impactent plusieurs *TerritoryVersion*, *LevelVersion* ou *UnitVersion* simultanément. Il offre ainsi une

compréhension globale du changement opéré. Des sous-classes spécialisent ce concept général de changement.

Une description détaillée de chaque concept (dont un positionnement par rapport à l'état de l'art) est donnée à l'adresse <http://purl.org/net/tsnchange>. Nous discutons ci-après de la question de l'*Identité* d'une unité territoriale qui conditionne une utilisation adaptée des concepts de l'ontologie TSN-Change.

2.3.1.2 Importance de l'Identité dans la détermination des changements

L'utilisation des ontologies *TSN* et *TSN-Change* requiert de définir ce qui détermine l'*Identité* d'un concept *UnitVersion*⁵³ dans la nomenclature. Le plus souvent l'*Identité* d'une unité territoriale est un composant complexe qui se définit par une liste d'attributs pondérés intégrant aussi bien un code (ou *identifier*), que le nom ou encore la géométrie de l'unité⁵⁴. Cette *Identité*, quelle que soit sa définition, permet de distinguer les situations relevant des trois états principaux *création* (affectation d'une nouvelle *Identité*), *continuation* (conservation de l'*Identité*), *élimination* (suppression de l'*Identité*) attendus d'un modèle traitant le problème d'évolution de l'identité (Hornsby and Egenhofer, 2002). Nous exploitons ces notions dans la caractérisation des changements en étudiant les filiations entre unités territoriales.

Deux cas de filiation peuvent être observés entre deux versions d'unités territoriales: la *continuation* désigne une filiation dans laquelle l'*Identité* est préservée alors qu'une *dérivation* implique une modification de l'*Identité* (Del Mondo, 2011). Autrement dit, une *continuation* est une relation entre deux *UnitVersion* ayant la même *Identité* alors qu'une *dérivation* est une relation entre deux (ou plus) *UnitVersion* aux *Identités* distinctes. Dans ce dernier cas, les deux *UnitVersion* font donc référence à deux instances différentes du concept *Unit* (voir Figure 2). Dans la Figure 4, les cadres étiquetés *Continuation* et *Dérivation* (en bas à gauche) montrent que les sous-types de changement recensés par (Plumejeaud, 2011) pour les restructurations de type *Merge*, *Split* et *Redistribution* peuvent être classifiés selon cette vision : ainsi, ce qui distingue une *Integration* d'une *Fusion* c'est la persistance d'une unité (au sens où elle continue d'exister sous la même *Identité*) d'une version de nomenclature à une autre.

Considérons le cas simplifié à l'extrême dans lequel l'*Identité* d'une unité territoriale serait exclusivement définie par la valeur d'un attribut *Identifier*. Une illustration du changement *Split* est proposée en Figure 5 selon les situations : *Derivation* (a) le *Split* se traduit alors par une *Scission* car *TU1* n'existe plus après le changement (perte de l'*Identité*) ; *Continuation* (b), le *Split* est alors représenté par une *Extraction* car l'unité d'*Identité* *TU1* existe toujours.

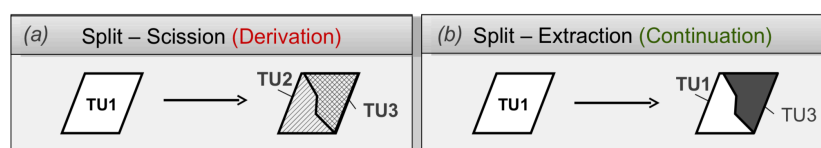


Figure 5 - Division d'une unité : distinction entre Scission et Extraction

Auteur: Camille Bernard, 2018.

2.3.1.3 Principes pour l'expression des changements

La Figure 6 vise à illustrer différentes situations que l'on peut rencontrer dans l'expression des changements entre versions de nomenclature. Pour plus de simplicité dans l'exposé, nous continuons d'assimiler l'*Identité* à la valeur de l'attribut *Identifier*. Les situations de filiation de type *Continuation* observée pour une unité donnée (donc dans le cas d'un *FeatureChange*) sont celles dans lesquelles on n'observe pas de changement de type *IdentifierChange* (i.e. maintien de l'*Identité*). Dans ce cas, le prédicat *hasNextVersion(x,y)* qui signifie "la *Version* *x* succède à la *Version* *y*" pourra être utilisé.

⁵³ Nous simplifions ici volontairement en restreignant le discours au cas des unités territoriales (i.e. *UnitVersion*) mais les territoires (*TerritoryVersion*) pourraient être concernés.

⁵⁴ Cette question sera ré-abordée plus en détail en section 2.4.1. Dans la suite *Identité* ainsi écrit fait référence à cette liste d'attributs dont les valeurs considérées conjointement permettent d'identifier une et seule unité territoriale.

C'est ce qui est illustré par les changements (a) et (b) de la Figure 6 : (a) montre un lien de filiation direct uniquement car l'unité reste porteuse de la même *Identité* symbolisée par *TU1*⁵⁵ (aucun autre attribut n'a changé par ailleurs) ; (b) illustre un lien de filiation similaire (on a toujours *TU1* comme *Identité*) accompagné de la description d'un changement (de type *Expansion*) d'un attribut non porteur de l'identité de l'objet, la géométrie de l'unité dans cet exemple⁵⁶, dont on observe l'augmentation de surface dans la nomenclature *TSN Version 3* par rapport à la précédente version). Dans le cas (b), aucun changement de type *StructureChange* n'apparaît : cela signifie qu'aucune autre unité territoriale existant dans la *TSN Version 2* n'est impliquée dans cette évolution⁵⁷. Cet exemple illustre qu'il est possible dans notre approche de représenter l'évolution territoriale résultant par exemple de l'annexion d'une unité située hors du territoire d'étude d'origine.

Dans des cas de *Dérivation* nous choisissons, contrairement à (Harbelot *et al.*, 2015), de ne pas créer de prédicats qui permettraient de lier directement une unité à une autre. En effet, dans bien des cas, ces liens risqueraient de ne pas être pertinents. Par exemple, dans des cas de redistribution, modifiant les frontières de plusieurs unités et leurs identifiants, il y a peu d'intérêt à établir un lien direct d'unité à unité, celui-ci n'étant que difficilement qualifiable et justifiable, de manière simple, d'un point de vue sémantique. Il nous paraît préférable de déterminer l'ensemble des unités impactées en entrée du changement et l'ensemble des unités créées en sortie de ce changement et de les intégrer dans une structure porteuse de sens. C'est en cela que nous nous revendiquons de l'approche *ChangeBridge* de (Kauppinen *et al.*, 2008; Kauppinen and Hyvönen, 2007) que nous jugeons plus adaptée au domaine des TSN et à la nécessité de décrire le changement. La Figure 6 (c) montre une relation de *dérivation* traduite par :

- un changement de type *StructureChange*, et plus précisément de type *Scission*, montrant que l'unité *TU1* se scinde en deux nouvelles unités *TU2* et *TU3*
- trois changements de type *FeatureChange*, chacun étant lié aux différentes unités mentionnées ci-dessus et traduisant ce que chacune subi comme événement. On note ainsi la disparition dans *TSN Version 4* de l'unité territoriale jusqu'ici identifiée par *TU1* : cela est traduit par le changement *Disappearance* dont *TU1* dans *TSN Version 3* est l'entrée. Les créations des unités *TU2* et *TU3* sont représentées par deux changements *Appearance* dont elles sont une sortie.

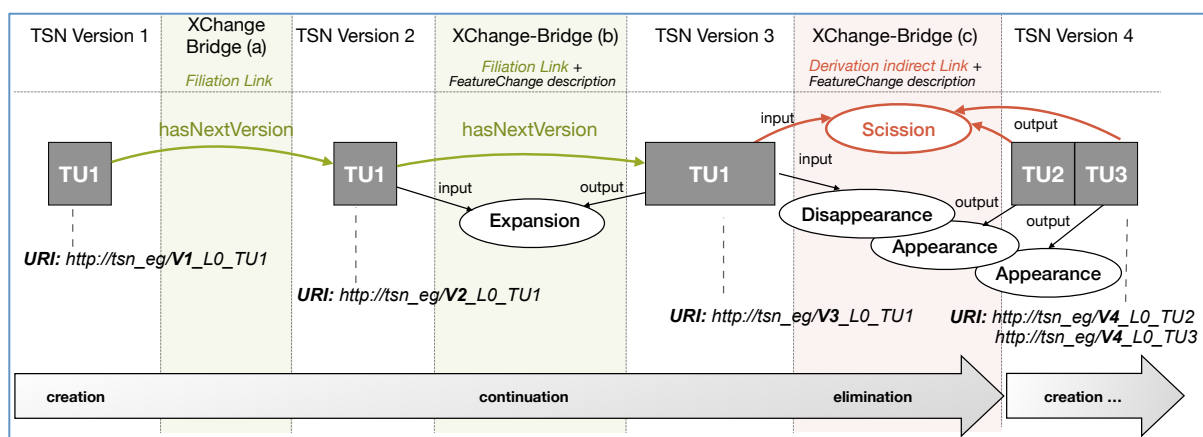


Figure 6 - Illustration de changements décrits à l'aide de TSN-Change Ontology

Auteur: Camille Bernard, 2018.

⁵⁵ Il convient de distinguer l'*Identifiant* (un code interne à la nomenclature) dont la valeur peut être la même pour deux *UnitVersion* car elles correspondent à une seule et même unité mais représentées à des temps différents, de l'URI que nous construisons pour chaque *UnitVersion* à des fins de publication dans le web des LOD conformément à l'Extrait de code 2.

⁵⁶ En réalité, l'*Identité* d'une unité territoriale est bien rarement considérée comme totalement indépendante de sa géométrie.

⁵⁷ Dans le cas contraire, nous aurions ici affaire à un *StructureChange* de type *Integration* car l'*Identifiant* de *TU1* resterait inchangé mais l'augmentation de surface serait due à l'absorption complète d'une autre unité, préexistante dans *TSN Version 1*, dans les (nouvelles) limites de *TU1*.

Les représentations des changements d'une version à une autre telles qu'illustrées dans Figure 6 désignent ce que nous nommons des *XChange-Bridge*, pour *eXtended Change Bridge*, en référence à l'approche (Kauppinen *et al.*, 2008; Kauppinen and Hyvönen, 2007). Le terme « *extended* » marque la différence de notre proposition qui s'appuie sur une modélisation étendue (*i.e.* plus complète) d'un changement comme expliqué dans la section suivante.

2.3.1.4 Expression multi-niveau des changements

L'ontologie *TSN-Change Ontology* permet la description de graphes des changements territoriaux multi-niveaux (voir Figure 7). Les changements sont décrits selon les principes exposés précédemment pour des composants d'une nomenclature spécifiés en entrée (prédicats *input* et son inverse *before*) et/ou en sortie (prédicats *output* et son inverse *after*). Le changement ainsi décrit concerne généralement un niveau de la TSN et peut être relié à des changements ayant lieu plus bas ou plus haut dans la hiérarchie TSN, s'il fait sens de relier ces changements. Cela est porté par des prédicats *upperChange* et/ou *lowerChange*.

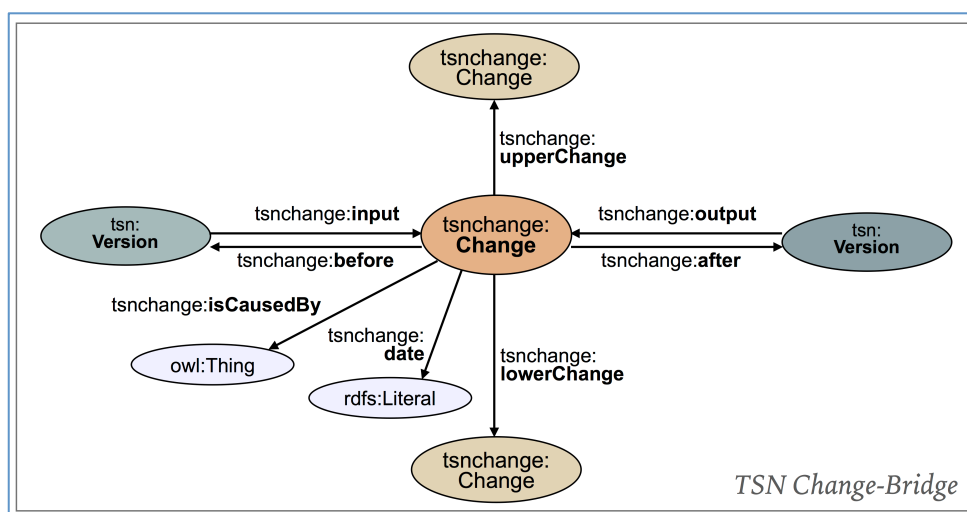


Figure 7 - Représentation d'un changement dans TSN-Change Ontology

Source : <http://purl.org/net/tsnchange>, Auteur : Camille Bernard, 2018.

Nous reprenons ici l'approche développée dans (Plumejeaud *et al.*, 2011) qui considère qu'un changement est rarement isolé et indépendant des autres changements qui se produisent simultanément dans une zone donnée et l'étendons. On trouve ainsi associé à la définition d'un *Change* le tuple de degré 6 suivant :

$\langle \text{input, output, lowerChange, upperChange, date, isCausedBy} \rangle$

où *t* est un prédicat qui permet de dater le changement ; *isCausedBy* est un prédicat qui vise à donner du sens au changement par la référence à un événement législatif ou sociétal par exemple ; *input* est l'ensemble des ressources impactées par le changement ; *output* est l'ensemble des ressources créées ou modifiées suite au changement ; *lowerChange* est l'ensemble des changements causés par le changement *Change* actuellement décrit et *upperChange* indique le changement *Change* supérieur qui a provoqué le changement actuellement décrit. Nous montrons ici comment il nous permet de dissocier les deux sous-types du changement *Split* (*i.e.* *Extraction* et *Scission*, voir Figure 4). La *Scission* de l'unité dont l'Identifieur est *TU1* dans l'encadré (a) de la Figure 5 est définie par :

$\text{Scission}(\text{TU1}) \equiv \langle \{\text{TU1}\}, \{\text{TU2}, \text{TU3}\}, \{\text{Disappearance}(\text{TU1}), \text{Appearance}(\text{TU2}), \text{Appearance}(\text{TU3})\}, *, \emptyset, \emptyset \rangle$

où $*$ représente l'ensemble des changements (non décrits ici) ayant pour sous-changement le changement *Scission* ici décrit, aucune date ni référence à un événement n'est mentionné, et :

$\text{Disappearance}(\text{TU1}) \equiv \langle \{\text{TU1}\}, \emptyset, \emptyset, \text{Scission}(\text{TU1}), \emptyset, \emptyset \rangle$

$\text{Appearance}(\text{TU2}) \equiv \langle \emptyset, \{\text{TU2}\}, \emptyset, \text{Scission}(\text{TU1}), \emptyset, \emptyset \rangle$

$\text{Appearance}(\text{TU3}) \equiv \langle \emptyset, \{\text{TU3}\}, \emptyset, \text{Scission}(\text{TU1}), \emptyset, \emptyset \rangle$

Le changement de type *Extraction* modifiant l'unité dont l'*Identifieur* est *TU1* dans l'encadré (b) de la Figure 5 est défini par le tuple :

$$\text{Extraction}(\text{TU1}) \equiv \langle \{\text{TU1}\}, \{\text{TU1}, \text{TU3}\}, \{\text{Contraction}(\text{TU1}), \text{Appearance}(\text{TU3})\}, *, \emptyset, \emptyset \rangle$$

où :

$$\text{Contraction}(\text{TU1}) \equiv \langle \{\text{TU1}\}, \{\text{TU1}\}, \emptyset, \text{Extraction}(\text{TU1}), \emptyset, \emptyset \rangle$$

$$\text{Appearance}(\text{TU3}) \equiv \langle \emptyset, \{\text{TU3}\}, \emptyset, \text{Extraction}(\text{TU1}), \emptyset, \emptyset \rangle$$

2.3.1.5 Gestion de l'évolution des frontières d'un territoire d'étude

Notre approche permet de décrire les changements affectant les frontières extérieures du territoire d'étude, faisant évoluer en cela le travail initial de (Plumejeaud et al., 2011b). Ces nouvelles étiquettes (apparaissant sous un nœud *TerritoryChange*, voir Figure 4) permettent en effet de dissocier les événements survenus dans un territoire dont les limites n'ont pas changé, des événements qui modifient l'étendue spatiale de la zone d'étude. Prenons l'exemple d'une fusion de deux territoires qui entraînerait une modification des unités territoriales, existant avant la fusion, et ce, sur plusieurs niveaux de la TSN comme illustré par la Figure 8. Dans cette figure, le territoire *A* couvert par la nomenclature *TSN Version 1* apparaît en vert et le territoire *B*, non couvert par cette nomenclature est symbolisé en gris.

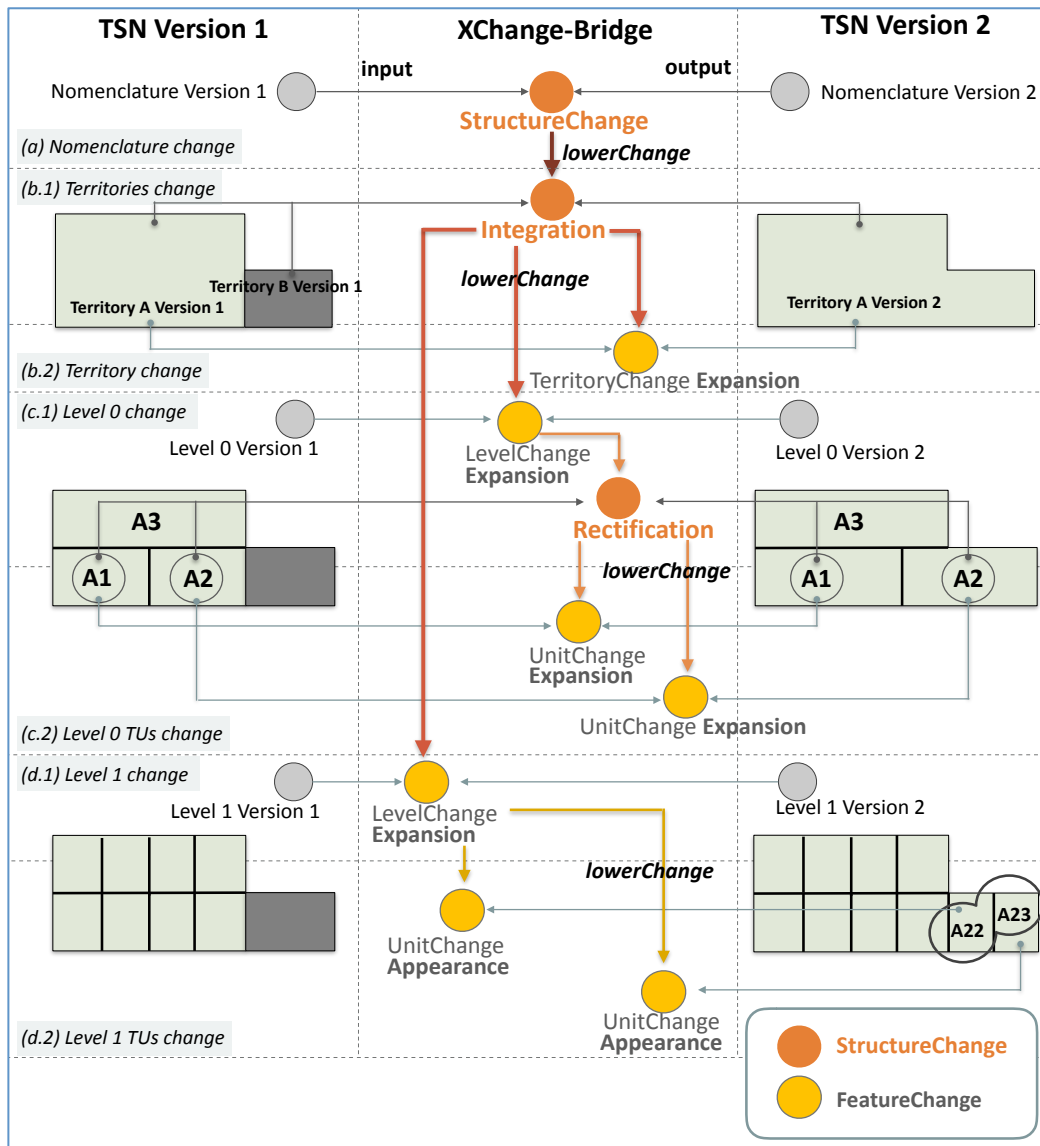


Figure 8 - Graphe multi-niveaux simplifié d'une fusion de deux territoires fictifs

Représentation selon les termes de TSN-Change Ontology, Auteur: Camille Bernard, 2018.

(a) Le nœud père d'un graphe du changement est toujours un nœud de type *StructureChange* qui indique quelles sont les versions de la TSN avant et après le changement décrit. Ce nœud père constitue un point d'entrée dans le graphe d'un changement territorial, bien souvent constitué de sous changements. Ensemble, ces nœuds construisent ce que nous nommons, un *graphe du changement territorial multi-niveaux* retraçant les répercussions qu'un changement affectant un élément père a sur ses sous-éléments. Les changements apparaissent en cascade, du fait de la structure hiérarchique des TSN. (b.1) Le nœud père est spécialisé par un nœud de type *Integration*⁵⁸, au premier niveau géographique de la TSN que constitue le *Territoire* d'étude. Cette étiquette indique que plusieurs objets géographiques – ici les territoires *Territoire A* et *B* en *input* du changement – fusionnent mais qu'un des territoires, le *Territoire A* en *output*, garde son identité suite au changement (*i.e.*, il s'agit d'un changement de type *Continuation*). Chacune des modifications provoquées par cette fusion est ensuite représentée, à chaque niveau de la TSN impacté. (b.2) L'impact sur l'objet *Territory A Version1* est d'abord décrit en terme d'*Expansion*. (c.1) Le niveau 0 de la nomenclature subit un changement (nœud *LevelChange*) entraînant une *Rectification* du tracé de ses unités. (d.1) Le niveau 1 de la nomenclature subit lui aussi un changement, mais contrairement au niveau 0, ses unités appartenant à la version 1, ne subissent pas de modification de frontières. Enfin, chaque changement subit par une unité est décrit. (c.2) Au niveau 0 de la nomenclature, cette fusion des territoires provoque une *Expansion* des unités *A1* et *A2*. (d.2) Au niveau 1, la fusion des territoires engendre la création des UT *A22* et *A23* (*Appearance*).

Dans la section suivante, nous revenons à l'exemple de la NUTS pour illustrer sur un cas réel comment les changements de version à version sont décrits.

2.3.2 Cas d'étude : Description des changements dans la NUTS

L'exemple des deux enclaves espagnoles en Afrique du Nord, nommées *Ceuta* et *Melilla* est considéré ici. Distantes spatialement, elles sont considérées comme une seule unité territoriale dans la version 1999 de la NUTS avec comme code ES63. La géométrie multi-polygone de cette unité est présentée en rouge, sur la carte de localisation à gauche de la Figure 9. Cette figure représente, de manière simplifiée le graphe RDF du changement territorial multi-niveaux de l'unité ES63 se scindant en deux entre les versions 1999 et 2003 de la NUTS. Le vocabulaire des ontologies TSN et TSN-Change est utilisé. Dans le cadre (a) est représenté le nœud père du changement (de type *StructureChange*) qui entraîne deux autres changements reliés par des prédicats *tsnchange:lowerChange* : une *Extraction* au Niveau 2, représentée dans le cadre (b), et un changement de type *IdentificationRestructuration* au Niveau 3 dont la représentation est déportée dans le cadre (c). Notez la flèche discontinue associée au prédicat *tsnchange:lowerChange* qui part du cadre (a) et se termine dans le cadre (c). Nous détaillons à présent les différentes parties de ce graphe du changement.

(a) tout en haut du graphe, le nœud père du changement est un nœud de type *StructureChange* dont l'URI est http://purl.org/steamer/nuts/change_nomenclaturechange_1999_2003⁵⁹. La ressource *NUTS Version 1999* de type *NomenclatureVersion* est un objet en entrée (prédicat *input*) de ce changement, et la ressource *NUTS Version 2003* est un objet en sortie de ce changement (prédicat *output*). Dans notre approche, tout nœud père d'un graphe décrivant un changement territorial est relié en entrée et en sortie à des ressources de type *NomenclatureVersion* (l'élément versionné le plus élevé dans la structure hiérarchique d'une TSN) : c'est ainsi qu'une machine peut reconnaître le point d'entrée dans un *graphe du changement territorial multi-niveaux*.

(b.1) le nœud père du changement est associé par le prédicat *lowerChange* à un changement de type *Extraction* observé à un niveau inférieur de la NUTS (au *Level 2*). Nous parlons ici d'extraction car ni l'identité du niveau, ni celle de l'unité territoriale en *input* ne sont modifiées entre la version NUTS 1999 et la version 2003 (*i.e.*, ils gardent respectivement leur *Identifieur* dans la NUTS, à savoir L2 et ES63, après le changement) : il s'agit donc d'une situation de *Continuation* qui nous permet de préciser la nature de la scission.

⁵⁸ Voir Figure 4 pour les différents types de changement mentionnés dans ce paragraphe.

⁵⁹ Selon une approche similaire à celle présentée dans l'Extrait de code 2, un patron pour les URIs du changement a été proposé, non détaillé ici.

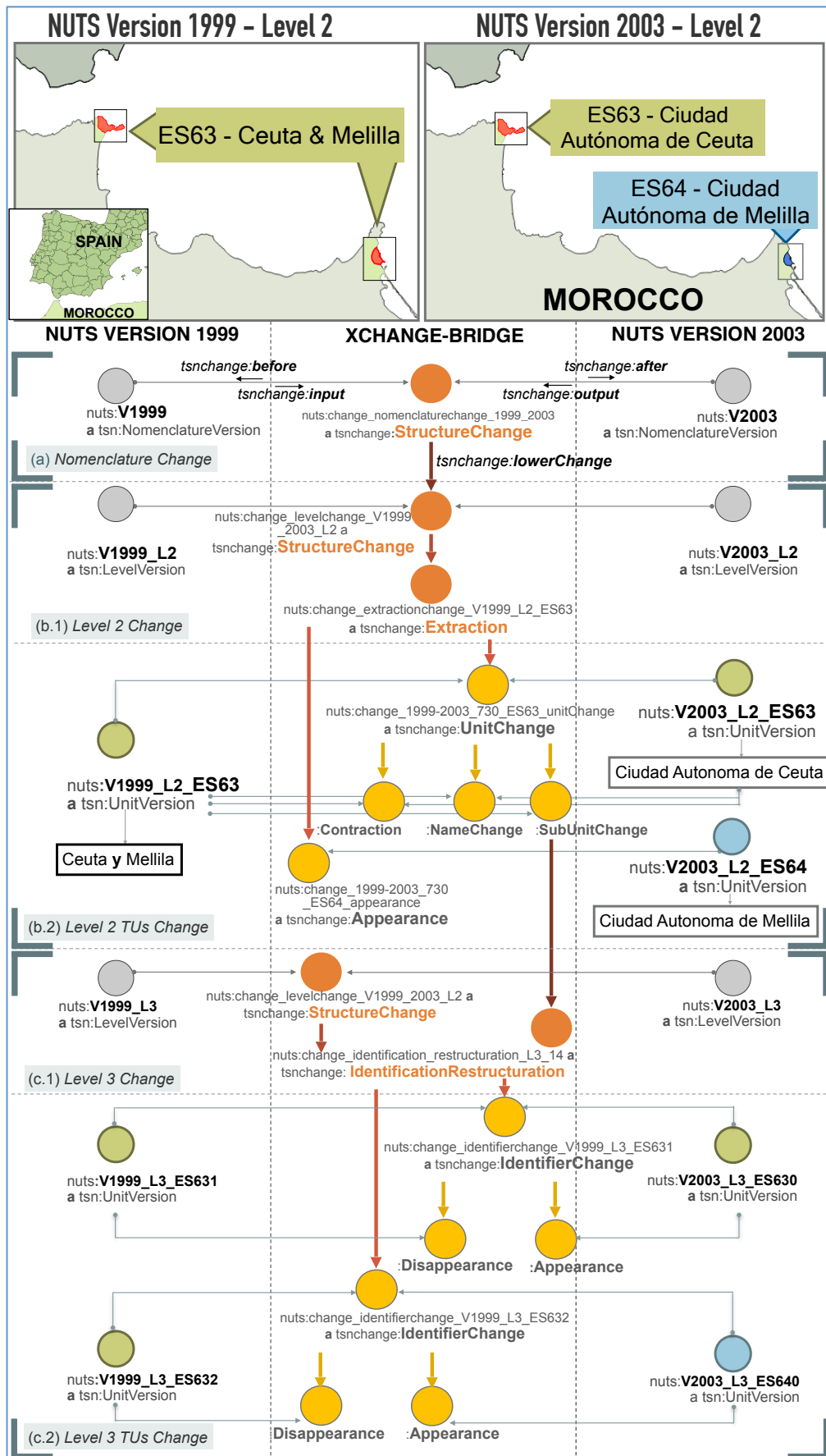


Figure 9 - Graphe multi-niveaux simplifié de la scission de l'unité territoriale ES63

Représentation selon les termes des ontologies TSN et TSN-Change, Auteur: Camille Bernard, 2018.

(b.2) les modifications subies par chaque unité du *Level 2* et dues à ce changement *Extraction* sont représentées via des nœuds de type *FeatureChange*. Ici, la scission de l'unité ES63 a pour conséquence, d'une part, un changement de type *UnitChange* qui lie la *Contraction* de l'unité ES63 et son changement de nom (nœud *NameChange*) au profit de « *Ciudad Autónoma de Ceuta* », et, d'autre part, la création (nœud *Appearance*) de l'unité ES64 (nommée « *Ciudad Autónoma de Melilla* »).

(c.1) la scission de l'unité ES63 a également des conséquences sur les sous-unités de ES63, au *Level 3* de la NUTS (ces sous-unités ne sont pas représentées sur la carte de la Figure 9, mais apparaissent dans le graphe). Une modification de l'ensemble des identifiants de nomenclature des sous-unités est observée. Un nœud du changement de type *IdentificationRestructuration* est donc créé et relie les deux *LevelVersion* (*V1999_L3* et *V2003_L3*) concernés pour consigner ce changement au niveau 3 de la nomenclature.

(c.2) sont ensuite détaillés les changements subis par chaque unité du niveau *Level 3* : les unités ES631 et ES632 changent toutes deux d'identifiants (nœuds *IdentifierChange*). L'identifiant de l'unité ES631 devient ES630 et l'identifiant de l'unité ES632 devient ES640⁶⁰. Dans la NUTS, l'attribut *identifier* d'une unité compte très fortement dans la constitution de l'*Identité*⁶¹, donc tout changement de valeur est susceptible de provoquer la disparition de l'unité qui le portait et la création d'une nouvelle unité (et ce, d'autant plus si les géométries ont changé, ce qui est le cas ici). Nous avons ici affaire à une situation de *Derivation* qui explique les deux sous changements portés par les nœuds *Disappearance* et *Appearance*. Il faut ici noter que grâce au nœud *IdentifierChange*, le linéage entre ces deux unités portant des identifiants différents est préservé, permettant ainsi de relier ces deux unités dans le temps.

Nous avons montré comment notre approche permet de décrire de façon détaillée les changements territoriaux observables d'une version de nomenclature à une autre. Grâce aux ontologies TSN et TSN-Change, nous sommes en mesure de déverser ces connaissances dans le Web des LOD. L'étape suivante à laquelle nous nous sommes consacrés est celle de l'automatisation de la création des graphes RDF correspondant aux changements. Dans ce but nous avons adapté l'*Algorithme d'appariement automatique de deux versions de nomenclature* décrit dans (Plumejeaud, 2011).

2.4 Algorithme de détection des changements entre versions de TSN

L'algorithme que nous proposons est décrit dans (Bernard et al., 2018a). Nous en donnons ici simplement les grands principes en nous appuyant sur une présentation simplifiée de ses étapes (voir Extrait de code 3). L'objectif de cet algorithme est de produire un graphe RDF des changements, conformes à l'ontologie TSN-Change, observés entre deux versions successives notées *V'* et *V''* d'une nomenclature. Adapté de (Plumejeaud, 2011) pour montrer la faisabilité de l'approche, l'algorithme actuellement développé prend en entrées deux listes (une pour *V'* et une pour *V''*) d'unités territoriales au format Postgis (et non à ce jour des instances d'ontologie TSN). Ces listes sont obtenues grâce à l'utilisation de l'outil *Statistical Units Nomenclature Integrator* (Bernard et al., 2017b) que nous avons mis au point dans le cadre du projet européen M4D⁶². Elles contiennent a minima pour chaque unité territoriale de la nomenclature (NUTS en l'occurrence) : son identifiant, son nom, son niveau d'appartenance dans la nomenclature, l'identifiant de sa super-unité, sa géométrie. L'approche adoptée est de comparer les unités pour déterminer leurs similarités et leurs différences et décider des liens de filiation à établir (*i.e.* créer les prédicats et nœuds de changements adaptés). Des parcours niveau par niveau de la nomenclature sont réalisés successivement, en partant du plus haut niveau (celui du Territoire) jusqu'au plus bas (celui des plus petites unités dans la hiérarchie que définit la nomenclature). Des matrices sont utilisées pour faciliter ces parcours et pour construire et stocker temporairement la connaissance relative aux correspondances et aux changements trouvés avant de produire le graphe RDF final. Les principales étapes sont :

⁶⁰ Une nouvelle convention semble être adoptée pour indiquer qu'une unité n'a qu'une seule sous-unité : dans ES630, le '0' semble indiquer que ES630 est la seule sous-unité de ES63 (elle aurait été numérotée ES631 sinon)

⁶¹ Ceci est présenté plus en détail dans la section 2.4.1.

⁶² Projet M4D (*Multi Dimensional Database Design and Development*), financé par le programme ESPON Database 2013 Phase II (2011-2014) <https://www.espon.eu/programme/projects/espon-2013/scientific-platform/espon-database-2013-phase-ii-2011-2014>. Voir Annexe 1, page 105 pour une description des travaux menés dans ce cadre.

- l'identification des changements entre versions
- la qualification sémantique des nœuds de changement
- l'établissement des liens entre nœuds de changement impactant des niveaux différents dans la nomenclature
- la génération du graphe des changements

```

Input: A list  $V'$  of  $m$  TUs and a list  $V''$  of  $n$  TUs, all TUs being at one Level  $l$ .
Output: RDF graph (RDFout) describing Matching and Changes on TUs
(Identifying changes)
for  $i$  in  $m$  do
  for  $j$  in  $n$  do
    SPATIALMATCH[ $i$ ][ $j$ ] ← SpatialMatchTest( $u'_i, u''_j$ );
    foreach  $u'_i.attribute$  do
      GLOBALMATCH[ $param.attribute$ ][ $i$ ][ $j$ ] ← DistanceTest( $u'_i.attribute, u''_j.attribute$ );
      (Also fills the SUPERMATCH $_l$  map at current level  $l$ )
    GLOBALMATCH[ $param.global$ ][ $i$ ][ $j$ ] ← WeightingTest( $u'_i, u''_j$ );

(Identifying clusters of changes)
for  $i$  in  $m$  do
  for  $j$  in  $n$  do
    (Identifying each clusters)
    if SPATIALMATCH[ $i$ ][ $j$ ] == Intersection then
      STRUCTURECHANGE[ $i$ ][ $j$ ] ← FindStructureChange( $u'_i, u''_j$ );
      (Also refines SPATIALMATCH[ $i$ ][ $j$ ] type)

(Associating a meaning with clusters of changes)
for  $i$  in  $m$  do
  for  $j$  in  $n$  do
    if SPATIALMATCH[ $i$ ][ $j$ ] == StructureChange then
      if GLOBALMATCH[ $param.global$ ][ $i$ ][ $j$ ] == Match then
        SPATIALMATCH[ $i$ ][ $j$ ] ← RefineChangeType(
          Continuation, SPATIALMATCH[ $i$ ][ $j$ ]);
      else
        SPATIALMATCH[ $i$ ][ $j$ ] ← RefineChangeType(
          Derivation, SPATIALMATCH[ $i$ ][ $j$ ]);

(Creation of the RDFOUT file)
for  $i$  in  $m$  do
  for  $j$  in  $n$  do
    if GLOBALMATCH[ $param.global$ ][ $i$ ][ $j$ ] == Match then
      RDFOUT ← CreateTUGenealogyLink( $u'_i, u''_j$ );
    foreach  $u'_i.attribute$  do
      RDFOUT ← CreateChangeNode(GLOBALMATCH[ $param.attribute$ ][ $i$ ][ $j$ ]);
      RDFOUT ← ChainingChangeNode(GLOBALMATCH[ $param.attribute$ ][ $i$ ][ $j$ ], SUPERMATCH[ $l-1$ ].GetDataForFey( $u'_i.super$ ));
      ( $l-1$  means level up from  $l$ )
    if STRUCTURECHANGE[ $i$ ][ $j$ ] > 0 then
      idChange ← STRUCTURECHANGE[ $i$ ][ $j$ ];
      while STRUCTURECHANGE[ $i$ ][ $j$ ] == idChange do
        RDFOUT ← CreateChangeNode(SPATIALMATCH[ $i$ ][ $j$ ]);
        RDFOUT ← ChainingChangeNode(SPATIALMATCH[ $i$ ][ $j$ ],
          SUPERMATCH[ $l-1$ ].GetDataForFey( $u'_i.super$ ));

```

Extrait de code 3 - Vision simplifiée de l'algorithme de détection des changements entre versions de TSN

2.4.1 Identifier des changements entre versions de territoires

Pour identifier des changements entre versions, l'algorithme utilise le concept d'Identité d'une unité territoriale selon une approche comparable à (Wang *et al.*, 2005) : si une unité u' de V' est jugée comme étant similaire à une unité u'' de V'' sur la base de ce qui fait leurs identités, alors un lien de filiation entre elles doit être établi. Dans notre approche, ce lien de filiation peut traduire que les unités u' et u'' sont les mêmes (et, dans ce cas, le recours à un unique prédicat *hasNextVersion*(u', u'') permettra d'établir un lien direct) ou qu'elles présentent des différences, à des degrés divers et pour des attributs divers, qui justifient la création d'un nœud du changement traduisant ces différences. Il

s'agit donc de déterminer le degré de similarité entre les unités, en laissant toute la souplesse nécessaire pour exprimer ce qui fait l'identité d'une entité (*i.e.* quel(s) attribut(s)) et pour décider de seuil au delà duquel on considère qu'il ne s'agit plus exactement de la même unité. À ces fins, l'algorithme prend en paramètres la liste pondérée des attributs participants à la définition de l'identité et un seuil ϵ en deçà duquel on ne considèrera pas deux unités comme étant les mêmes.

Des opérateurs permettant de comparer deux à deux les attributs de u' et u'' issus de cette liste sont définis. Il s'agit de mesurer la distance qui sépare u' et u'' (actuellement la distance de Levenshtein (Levenshtein, 1966) est utilisée pour les attributs de type texte tels que le nom de l'entité⁶³, et la distance surfacique pour les géométries⁶⁴ (Bel Adj Ali & Vauglin, 1999)) et de restituer cette information de manière normalisée par une valeur comprise entre 1 (ressemblance totale) et 0 (dissemblance totale). Les fonctions *DistanceTest* et *SpatialMatchTest* sont utilisées pour les calculs de distance inter-attributs selon leur type, et la fonction *WeightingTest* permet de calculer le score de similarité global entre u' et u'' . L'Équation 1 (cadre supérieur) donne la spécification de la fonction *WeightingTest* (désignée par F). Le cadre inférieur illustre un cas de mesure de similarité basé sur une identité qui serait décrite (par un expert du domaine) comme étant composée de quatre attributs *identifier*, *geom*, *name*, *super* (pour super-unité) respectivement associés aux poids 1/3, 1/3, 1/6 et 1/6.

$$\begin{array}{l}
 F((\alpha_1, C_1), (\alpha_2, C_2), \dots, (\alpha_r, C_r)) \text{ where } \sum_r \alpha_k = 1 \\
 F((\alpha_1, C_1), (\alpha_2, C_2), \dots, (\alpha_r, C_r)) : \\
 (V' \times V'') \rightarrow \mathbb{R}, (v', v'') \mapsto p, p \in [0, 1], \forall v' \in V' \text{ and } \forall v'' \in V'' \\
 \\
 F(\alpha_1, \text{identifier}), (\alpha_2, \text{geom}), (\alpha_3, \text{name}), (\alpha_4, \text{super}) \\
 \text{ where } \alpha_1 = 1/3, \alpha_2 = 1/3, \alpha_3 = 1/6, \alpha_4 = 1/6 \\
 F(\alpha_1, C_1), (\alpha_2, C_2), \dots, (\alpha_r, C_r) : \\
 (V' \times V'') \rightarrow \mathbb{R}, \\
 (v', v'') \mapsto 1/3 * (|u'.\text{identifier} - u''.\text{identifier}|) \\
 + 1/3 * (|u'.\text{geom} - u''.\text{geom}|) + 1/6 * (|u'.\text{name} - u''.\text{name}|) \\
 + 1/6 * (|u'.\text{super} - u''.\text{super}|)
 \end{array}$$

Équation 1 – Calcul de similarité entre deux unités appartenant à des versions différentes de TSN

Source : (Bernard et al., 2018a)

À l'issue de la première double itération (voir Extrait de code 3, première boucle 'For'), deux matrices de n lignes et m colonnes (correspondants respectivement aux unités u' de V' et u'' de V'') sont complétées. La matrice *SpatialMatch* contient les valeurs des distances surfaciques calculées entre les géométries (valant 0 si les unités sont disjointes). La matrice *GlobalMatch* contient le score de similarité global calculé pour chaque couple d'unités non disjointe, 0 sinon. À l'issue de cette première étape, les unités qui n'ont pas changé, en tenant compte de la valeur du seuil ϵ fixé, sont connues : elles sont reliées par le prédicat *hasNextVersion*. On est aussi en mesure de déterminer⁶⁵ les

⁶³ Nous prévoyons de revoir ce point à la lumière des résultats récents de (Santos et al., 2018) proposant une étude comparative des performances de différentes mesures de similarités appliquées à une tâche de *matching* de toponymes.

⁶⁴ Cette distance calcule le ratio entre l'aire partagée par deux géométries qui s'intersectent et l'union de leurs aires. Quand les géométries sont égales ou presque, ce ratio est proche de 1, sinon il tend vers 0. Un seuil β est utilisé pour définir la proportion d'aire que doivent partager à minima deux géométries pour être considérées comme identiques. Utilisée sur des géométries au préalable transformées dans le même Système de Référence de Coordonnées, cette distance permet de les comparer même si elles n'ont pas été générées avec le même niveau de détail (ce que ne permet pas un test tel que *ST_equals* proposé par l'OGC)

⁶⁵ Cette étape repose sur l'analyse des données des matrices. Un cas trivial est celui de la disparition : dans les matrices, la ligne de l'unité u' de V' n'a que des scores à 0 ou en dessous du seuil ϵ pour toutes les colonnes correspondant aux unités u''

unités qui ont subi un changement, qui sont apparues ou qui ont disparu et de leur associer le *FeatureChange* adapté.

L'étape suivante de l'algorithme est de déterminer les changements qu'il convient de regrouper pour créer les nœuds de type *StructureChange*.

2.4.2 Regrouper et qualifier les changements

Un second parcours de la matrice *SpatialMatch* permet à la fonction *FindStructureChange* de constituer des ensembles d'unités impliquées dans un même *StructureChange*. L'identification d'un *StructureChange* repose sur la recherche d'un invariant : « l'union des géométries préalablement calculée avant un changement K est égal (ou presque égal selon la distance surfacique) à l'union des géométries après ce changement K ». Deux listes d'unités sont construites et l'union de leurs géométries comparées : la liste *beforeK* recense des unités u' de V' avant le changement, la liste *afterK* recense des unités u'' de V'' après le changement. La liste *beforeK* est initialisée avec une unité u' de V' dont on sait qu'elle est en intersection avec au moins une unité u'' de V'' (grâce aux valeurs de la matrice *SpatialMatch*). Ces unités u'' de V'' sont ajoutées à la liste *afterK*. Partant cette fois de la liste des unités *afterK*, on ajoute à la liste *beforeK* toute unité de V' qui intersecte ce nouvel ensemble d'unité. La procédure se poursuit jusqu'à ce qu'au moins une des listes *beforeK* ou *afterK* reste identique par rapport à son état précédent.

À l'issue du processus, les nœuds de type *StructureChange* sont identifiés et pour chacun on connaît la liste des unités en entrée (celles de *beforeK*) et en sortie (celles de *afterK*). Il est également possible de préciser le type de *StructureChange* (voir Figure 4). En effet :

- si la liste *beforeK* ne contient qu'une seule entité, alors on est en présence d'un *Split*
- si la liste *afterK* ne contient qu'une entité, alors on a affaire à un *Merge*
- dans les autres cas, il s'agit d'une *Redistribution*.

La spécification encore plus précise du type de changement pour ces trois sous-types dépend du fait d'avoir affaire à une situation de *Continuation* ou de *Dérivation*. Cela est déterminé dans la fonction *RefineChangeType* par l'analyse des *FeatureChange* associés à l'étape précédente aux unités impliquées dans chaque *StructureChange*. Si au moins une unité en entrée du *StructureChange* conserve son *Identité* entre les deux versions étudiées alors il s'agit d'un cas de *Continuation*.

2.4.3 Relier les changements

La dernière étape de l'algorithme consiste à relier les changements qui concernent des niveaux différents dans la nomenclature mais sont connectés pour enrichir encore la description de l'évolution territoriale. Ce chainage permet des parcours *top-down* et *bottom-up* lors de l'exploration des changements, offrant ainsi une vision de leurs origines ou de leurs répercussions à tous les niveaux de la structure hiérarchique.

Chaque changement va ainsi être connecté aux changements de niveau supérieur et/ou inférieur, s'ils existent, via les prédicats *upperChange* et *lowerChange*, et ce pour tous les niveaux de la hiérarchie. Nous exploitons pour cela les relations structurelles entre les composants d'une nomenclature telle que l'appartenance d'une unité à un niveau (prédicat *isMemberOf*) ou encore la relation hiérarchique entre une unité et sa super-unité (prédicat *hasSuperFeature*) définies dans l'ontologie TSN. Une double approche ascendante et descendante est menée à partir des changements identifiés aux étapes précédentes pour définir correctement le chainage. Nous renvoyons le lecteur à (Bernard et al., 2018a) pour une présentation détaillée des étapes de l'algorithme sur ce point.

2.4.4 Produire le graphe RDF

La création du graphe RDF est accomplie en fin de traitement, à partir des informations contenues dans les différentes matrices construites et complétées au fur et à mesure (voir Extrait de code 3, dernière double boucle 'For'). Les opérations suivantes sont réalisées :

de V'' . Le cas de l'apparition est une transposée de cette explication. Les autres changements supposent d'observer les valeurs pour chaque attribut.

1. le versionnement de toutes les ressources pour une immersion dans le web des LOD en créant pour chacune un identifiant unique conformément aux *patterns* que nous avons proposés (voir Extrait de code 2)
2. la création de liens directs entre les versions d'unités dont l'identifiant ne change pas de V' à V'' à l'aide du prédicat *tsnchange:hasNextVersion* (fonction *CreateTUGenealogyLink*).
3. la création d'un nœud du Changement (fonction *CreateChangeNode*) selon la typologie de TSN-Change Ontology et en décrivant les entrées et sorties avec les prédicats correspondants.
4. le chainage inter-niveaux des changements à l'aide des prédicats *tsnchange:upperChange* et *tsnchange:lowerChange* (fonction *ChainingChangeNode*).

2.4.5 Cas d'étude : détection de changements dans la NUTS

L'algorithme que nous avons élaboré a été testé sur notre cas d'étude, pour les versions NUTS 1999 à 2010. Une implémentation Java a été réalisée, incluant l'API Apache-Jena RDF pour créer le graphe de sortie. Dans le test réalisé, les attributs pris en compte pour établir la similarité entre unités sont le code de l'unité (*identifier*), la géométrie (*geom*), le nom (*name*), et le code de l'unité supérieure (*super*). Les poids assignés à ces attributs sont les suivants: $\alpha_{\text{identifier}} = 0.4$, $\alpha_{\text{geom}} = 0.4$, $\alpha_{\text{name}} = 0.1$, $\alpha_{\text{super}} = 0.1$. Le seuil ϵ de détermination d'une correspondance identitaire entre unités est fixé à 0,005.

Le programme crée le graphe RDF entre deux versions dans un temps d'exécution compris entre 15 et 20 minutes⁶⁶ selon les cas. Chaque version de la NUTS compte environ 1800 unités territoriales, avec par exemple dans la NUTS 2016 104 unités au niveau NUTS 1, 281 unités au niveau NUTS 2 et 1348 unités au niveau NUTS 3.

Plus de 39306 ressources pour les quatre versions de nomenclatures étudiées (décrivant des correspondances directes et des changements) sont générées et publiées en ligne à partir d'un *triplestore* accessible à <http://purl.org/steamer/nuts>. Si l'on s'intéresse à la nature des 588 changements identifiés entre les versions NUTS 2006 et 2010, on compte 524 *FeatureChange* et 64 *StructureChanges* qui se répartissent en 8 *Reallocation*, 14 *Merge*, 9 *Scission*, 33 *IdentificationRestructuration*. Le nombre de *FeatureChange* est sensiblement plus élevé que les changements décrits dans les listes fournies par Eurostat⁶⁷ pour accompagner la publication d'une nouvelle version de nomenclature, mais cela s'explique puisque notre représentation des changements s'appuie sur une typologie plus détaillée et que nous construisons des nœuds de changements intermédiaires qui n'ont pas d'équivalents dans la vision Eurostat (par exemple les *SubUnitChange* et *SuperUnitChange*). Une autre forme d'évaluation de ces résultats peut être réalisée en adoptant un point de vue qualitatif sur la génération automatique des changements autorisée par notre algorithme :

- il retrouve tous les changements listés manuellement par Eurostat ;
- il donne une description plus fine des changements que celle fournie par Eurostat en s'appuyant sur notre typologie.
- il donne une vision inter-niveaux des changements non disponible jusqu'ici
- il peut tenir compte de la définition complexe de ce qui fait l'*Identité* d'une unité territoriale
- il est capable de détecter des changements impliquant de très petites unités du territoire
- il permet de détecter des cas particuliers comme un *Split* impliquant une unité multi-polygones dont les parties sont spatialement distantes. Le recours aux géométries uniquement permet d'identifier un *Split*, mais grâce à l'information hiérarchique, il est possible de savoir qu'une nouvelle unité territoriale, créée après le *Split*, est intégrée dans un nouveau territoire (i.e. une nouvelle unité mère). Il devient ainsi possible de décrire des cas d'annexion ou de cession de territoires. Cela est une conséquence du chainage inter-niveau qui fait que les changements subis par chaque partie de l'unité sont reliés à un changement au niveau supérieur (car l'algorithme tient compte de la hiérarchie des éléments et pas seulement de l'information spatiale) et qui peut alors être correctement interprété.

⁶⁶ Caractéristiques de la machine utilisée : iOS, 16GB de mémoire et processeur Intel Core i7 à 2,5 GHz.

⁶⁷ Voir par exemple <http://ec.europa.eu/eurostat/documents/345175/629341/1999-2003.xls>

Les graphes du changement produit par l'algorithme ont vocation à être mis à disposition dans le Web des LOD, depuis un SPARQL Endpoint et peuvent être exploités via le langage de requête SPARQL et son extension spatiale *GeoSPARQL* comme illustré dans la section suivante.

2.5 Exemple d'extractions de connaissances

2.5.1 Requêtage du SPARQL Endpoint

Nous illustrons ici par quelques exemples les possibilités offertes d'interrogation. Si l'énoncé d'une requête est assez simple et son exécution rapide, le résultat obtenu montre la richesse des connaissances portées par les ontologies *TSN* et *TSN-Change*.

Depuis le SPARQL Endpoint à l'adresse <http://purl.org/steamer/nuts> tous les triplets décrivant les points d'entrées dans les graphes du changement (entre deux versions de TSN) peuvent être extraits (voir Extrait de code 4) en filtrant les triplets pour n'obtenir que ceux dont le prédicat est *tsnchange:inputNomenclatureVersion*.

```
1 PREFIX tsnchange: <http://purl.org/net/tsnchange#> .
2 PREFIX nuts: <http://purl.org/steamer/nuts/> .
3 SELECT * WHERE { ?inputNomenclatureVersion
4     tsnchange:inputNomenclatureVersion ?parent_change_node .}
```

Extrait de code 4 - Requête SPARQL pour extraire les points d'entrées de graphes multi-niveaux du changement

Le résultat est illustré par l'Extrait de code 5. La liste de tous les points d'entrées des graphes du changement dans la NUTS est donnée : actuellement, trois point d'entrées sont disponibles, correspondants aux changements entre les versions 1999 et 2003, 2003 et 2006, et 2006 et 2010 de la NUTS.

```
1 nuts:V1999 tsnchange:inputNomenclatureVersion
2     nuts:change_nomenclaturechange_1999_2003 .
3 nuts:V2003 tsnchange:inputNomenclatureVersion
4     nuts:change_nomenclaturechange_2003_2006 .
5 nuts:V2006 tsnchange:inputNomenclatureVersion
6     nuts:change_nomenclaturechange_2006_2010 .
```

Extrait de code 5 - Extrait du résultat retourné par la requête SPARQL présentée dans l'Extrait de code 4

(Syntaxe RDF-Turtle)

À partir de cette réponse, il est possible d'explorer ensuite chacun des graphes du changement, entre deux versions de TSN, en utilisant l'opérateur SPARQL *DESCRIBE* appliqué successivement, sur les ressources retournées et pointées par le prédicat *tsnchange:lowerChange*. Ci-après (Extrait de code 6) la requête appliquée au nœud de changement entre les versions 1999 et 2003 retourne un résultat, affectant le niveau L2. La seconde requête (ligne 4) appliquée à ce nœud nous renseigne sur l'existence d'un changement de type *Extraction* impliquant l'unité ES63 décrite dans la version 1999.

```
1 DESCRIBE nuts:change_nomenclaturechange_1999_2003
2 # RESPONSE
3 nuts:change_nomenclaturechange_1999_2003 tsnchange:lowerChange
4     nuts:change_levelchange_V1999_2003_L2 .
4 DESCRIBE nuts:change_levelchange_V1999_2003_L2
5 # RESPONSE
6 nuts:change_levelchange_V1999_2003_L2 tsnchange:lowerChange
7     nuts:change_extractionchange_V1999_L2_ES63 .
```

Extrait de code 6 - Requêtes SPARQL DESCRIBE pour obtenir la description d'un nœud

Les triplets de l'Extrait de code 7 montrent comment, en suivant les nœuds indiqués par le prédicat *tsnchange:lowerChange*, il est possible de descendre la chaîne des changements, du nœud

Extraction de l'UT ES63 (entre les versions 1999 et 2003 de la NUTS), jusqu'aux nœuds décrivant les changements subis par ses sous-éléments, c'est-à-dire les unités ES631 et ES632. Le contenu présenté dans cet extrait est le même que celui présenté Figure 9.

```

1 nuts:change_extractionchange_V1999_L2_ES63 a tsnchange:Extraction;
2   tsnchange:unitVersionBefore nuts:V1999_L2_ES63 ;
3   tsnchange:unitVersionAfter nuts:V2003_L2_ES63, nuts:
4     V2003_L2_ES64;
5   tsnchange:upperChange nuts:change_levelchange_V1999_2003_L2;
6   tsnchange:lowerChange nuts:change_unitchange_V1999_L2_ES63;
7     nuts:change_appearance_V2003_L2_ES64.
8 nuts:change_unitchange_V1999_L2_ES63 a tsnchange:UnitChange;
9   tsnchange:unitVersionBefore nuts:V1999_L2_ES63;
10  tsnchange:upperChange
11    nuts:change_extractionchange_V1999_L2_ES63;
12  tsnchange:lowerChange nuts:change_namechange_V1999_L2_ES63, nuts
13    :change_geometrychange_V1999_L2_ES63, nuts:
14    change_subunitof_V1999_L2_ES63_change.
15 nuts:change_subunitof_V1999_L2_ES63_change a tsnchange:
16   SubUnitChange;
17   tsnchange:unitVersionBefore nuts:V1999_L3_ES632, nuts:
18     V1999_L3_ES631;
19   tsnchange:upperChange nuts:change_unitchange_V1999_L2_ES63;
20   tsnchange:lowerChange nuts:
21     change_identification_restructuration_L3_14.
22 nuts:change_identification_restructuration_L3_14 a tsnchange:
23   IdentificationRestructuration;
24   tsnchange:unitVersionAfter nuts:V2003_L3_ES630, nuts:
25     V2003_L3_ES640 ;
26   tsnchange:unitVersionBefore nuts:V1999_L3_ES631, nuts:
27     V1999_L3_ES632 ;
28   tsnchange:upperChange
29     nuts:change_subunitof_V1999_L2_ES63_change;
30   tsnchange:lowerChange nuts:
31     change_identifierchange_V1999_L3_ES632, nuts:
32     change_identifierchange_V1999_L3_ES631.
33 nuts:change_identifierchange_V1999_L3_ES631 a tsnchange:
34   IdentfierChange ;
35   tsnchange:unitVersionAfter nuts:V2003_L3_ES630 ;
36   tsnchange:unitVersionBefore nuts:V1999_L3_ES631 ;
37   tsnchange:upperChange nuts:
38     change_identification_restructuration_L3_14;
39   tsnchange:lowerChange nuts:change_appearance_V2003_L3_ES630,
40     nuts:change_disappearance_V1999_L3_ES631.

```

Extrait de code 7 – Extrait de la chaîne des changements de l'UT ES63 entre les versions NUTS 1999 et 2003

(Syntaxe RDF-Turtle)

2.5.2 Ouverture sur le Web des LOD

Nous pouvons enrichir les descriptions des TSN en recherchant d'autres ressources permettant de comprendre davantage le territoire observé. L'exemple trivial suivant montre comment utiliser le service DBpedia⁶⁸ pour découvrir notamment que l'unité ES63 est une Ville Autonome d'Espagne depuis 1995. L'interrogation se fait en concaténant au préfixe DBPedia, le nom de l'unité "*Ciudad Autónoma de Ceuta*" que nous pouvons obtenir par une requête sur nos graphes RDF portant sur le prédicat *hasName* pour l'unité considérée (http://purl.org/steamer/nuts/V2003_L2_ES63). L'URI ainsi

⁶⁸ <http://wiki.dbpedia.org/>

obtenue (http://dbpedia.org/resource/Ciudad_Autonoma_de_Ceuta) nous permet par exemple d'accéder à la population totale de cette unité (`dbo:populationTotal 82376^^xsd:integer`), information pouvant aider l'analyste au moment de comparer cette unité avec une autre. En particulier, le faible nombre d'habitants pour cette UT, en regard du nombre moyen d'habitants pour les unités de Niveau 2 de la NUTS (de 800000 à 3000000 en moyenne), devrait conduire les analystes à être prudents dans leurs explorations comparatives impliquant les données de cette unité. Il convient aussi de noter que cette valeur de population, retrouvée via DBPedia, ne peut être mobilisée qu'avec réserve dans des analyses expertes, car à défaut de métadonnées précises permettant de connaître la date de la mesure et les frontières précises de l'unité utilisées pour le relevé, il est impossible de juger de la qualité de ce nombre. Dans des analyses territoriales menées par des experts, il convient à terme de privilégier des données statistiques officielles dont les observations seront déclarées via l'ontologie statistique dédiée, *RDF Data Cube Vocabulary* du W3C (Cyganiak, Reynolds, 2014) interrogeable selon des principes similaires à ceux décrits ci-dessus. Nous revenons sur ce point dans la section suivante.

2.6 Synthèse et prolongements

Les recherches présentées dans ce chapitre sont au cœur de la thèse de Camille Bernard [TH-6]. Elles s'appuient sur et prolongent les contributions de (Plumejeaud, 2011) qui marquent les premiers travaux sur la représentation de l'évolution des territoires menés au sein de l'équipe STEAMER. Les motivations à l'origine de ces recherches sont, d'une part, d'accroître la qualité des descriptions des changements territoriaux en offrant des représentations plus justes, plus précises et plus complètes que l'existant et, d'autre part, d'actualiser les modalités de mise à disposition et d'exploitation des connaissances décrites en favorisant l'interopérabilité. Dès lors, le recours aux technologies du Web des LOD a été envisagé pour les possibilités offertes en regard de nos objectifs. Nos propositions se sont inscrites dans ce cadre en apportant des solutions sur différents points. Du côté des LOD, les spécifications RDF indiquent que le modèle de données est atemporel, mais que les graphes RDF peuvent décrire des événements et leur temporalité, via la définition de concepts ontologiques appropriés. Il est ainsi recommandé d'utiliser les vocabulaires existants (tels que OWL-Time) pour l'interopérabilité sémantique entre systèmes. Face aux manques de l'état de l'art concernant des vocabulaires dédiés à la représentation (telle que nous la souhaitons et qui font ses atouts) des évolutions territoriales, notre approche a été de proposer les ontologies TSN et TSN-Change. Leur élaboration a été guidée par une analyse des bonnes pratiques mais aussi de solutions à apporter à des questions conceptuelles et techniques. Nous avons par exemple suivi les recommandations du W3C (Cyganiak and Reynolds, 2014) indiquant qu'exprimer différents états d'une ressource au fil du temps peut se faire grâce à différents graphes RDF dédiés à ces états et inclus dans une ressource "contenant". Cela a conduit à la représentation d'une TSN via une ressource RDF, avec ses versions successives, elles-mêmes représentées par différents graphes RDF, ayant chacun un URI pour l'identification unique des états de la TSN au fil du temps. (Stefanidis *et al.*, 2014) par exemple relève que cette approche peut s'avérer lourde en termes d'espace de stockage des données. Une alternative consiste à ne stocker qu'une seule version (la première, le *version model* selon (Conradi and Westfechtel, 1998)) et les deltas (les différences entre deux versions consécutives). Néanmoins, cette solution requiert la génération des versions à chaque requête adressée aux systèmes, ce qui peut entraîner une surcharge, au moment du calcul cette fois-ci. Mais en termes d'usage, les différentes versions de TSN sont destinées à coexister dans le Web des LOD sous forme de documents directement consultables (*e.g.*, accéder à une version de NUTS en particulier). Ces versions ne constituent pas, par ailleurs, de grands ensembles de données qui évoluent rapidement (*e.g.*, la périodicité de la NUTS est d'environ trois ans). Ces aspects justifient l'approche que nous adoptons (*Materialization Using Versions and Deltas* (Stefanidis *et al.*, 2014)) qui consiste donc à conserver toutes les versions d'une même TSN et à décrire les deltas entre deux versions consécutives d'une TSN, car ces deltas, dans le contexte de la statistique, sont essentiels. Ces deltas sont loin d'être anecdotiques dans notre approche puisqu'ils constituent le concept de changement au cœur de l'ontologie *TSN-Change*. Par la description fine des changements territoriaux qu'elle autorise, cette ontologie alliée à l'ontologie *TSN*, offre un gain de qualité puisqu'il est désormais possible de référencer précisément une portion de territoire en particulier, c'est-à-dire une unité dans une version de nomenclature donnée, passée ou présente. L'extension *QB4ST* (Atkinson, 2017), pour une déclaration canonique de la dimension spatiale d'une observation statistique dans RDF Data Cube, est

une initiative à laquelle nous nous intéressons particulièrement pour promouvoir notre approche dans un cadre standardisé.

Pour faciliter le recours à nos contributions, des services web géographiques sont en cours de développement avec comme premier objectif d'exposer ces nouvelles connaissances sur l'évolution d'un territoire et d'apporter une aide à la décision dans leur utilisation. Ainsi, ces services (actuellement développés pour la NUTS) permettent, par exemple et entre autres, de découvrir qu'une unité territoriale n'existe pas en une seule version mais en plusieurs au gré de l'évolution des nomenclatures. Mieux informé, l'utilisateur peut alors décider de *la* version d'unité qu'il convient de lier à ses propres données (économiques, démographiques, sociologiques, environnementales, *etc.*) pour les territorialiser correctement. Cela nous ramène aux utilisateurs et aux besoins, évoqués en fin de section 2.1.1, auxquels nous ambitionnons d'apporter une réponse. Nous les rappelons et les commentons au vu de nos contributions qui nous permettent d'envisager des prolongements pour poursuivre ce travail.

a) *donner à voir les changements de frontières au cours du temps, en prendre conscience et ainsi mieux comprendre les dynamiques des territoires par l'évolution de leurs limites internes ou de celles de leurs voisinages.* Nous disposons à présent de capacités de représentations sémantiquement riches qui permettront de restituer visuellement les connaissances sur les changements territoriaux. Des sorties telles que celles que permet d'obtenir GraphDB sont un premier pas en ce sens. La Figure 10 à titre d'exemple montre une certaine stabilité de l'unité L2_ES63 à partir de la version NUTS 2003 (uniquement des prédicats *hasNextVersion*, alors qu'entre 1999 et 2003 des changements importants sont observés : l'*Extraction* (en jaune), le *nameChange* (en bleu) et le *GeometryChange* (en violet) qui correspondent aux changements décrits dans les cadres (b.1) et (b.2) de la Figure 9 (p26).

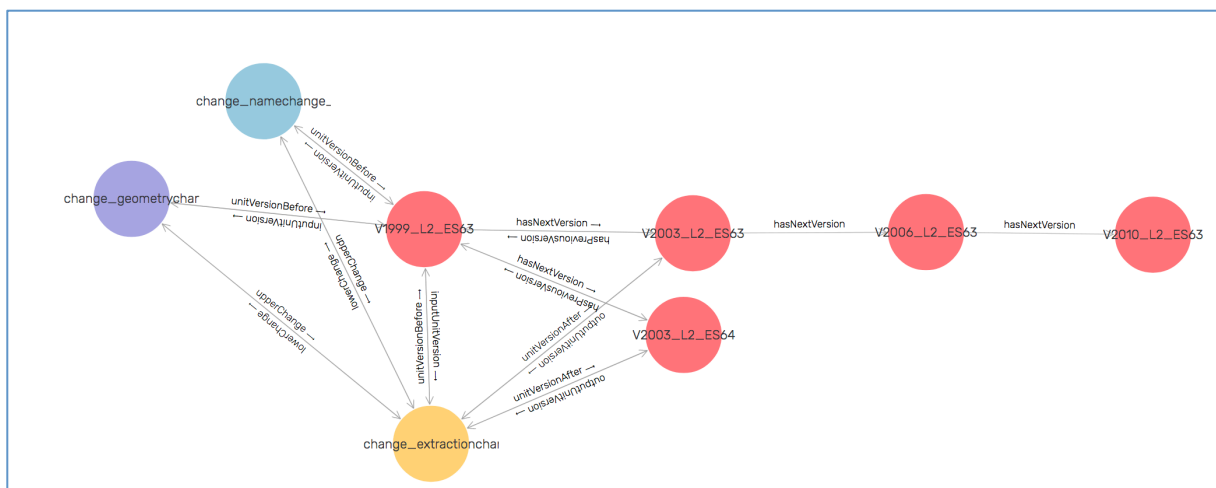


Figure 10 - Changements subis par l'unité L2_ES63 de la NUTS 1999 à la NUTS 2010

Visualisation sous GraphDB ; les nœuds sont interactifs (navigation dans le graphe en les développant), Auteur : C. Bernard, 2018.

Des restitutions cartographiques sont aussi à étudier et peuvent s'inspirer par exemple des sorties proposées par GitHub pour la visualisation des changements entre deux versions de fichiers au format *geojson* (voir Figure 11).

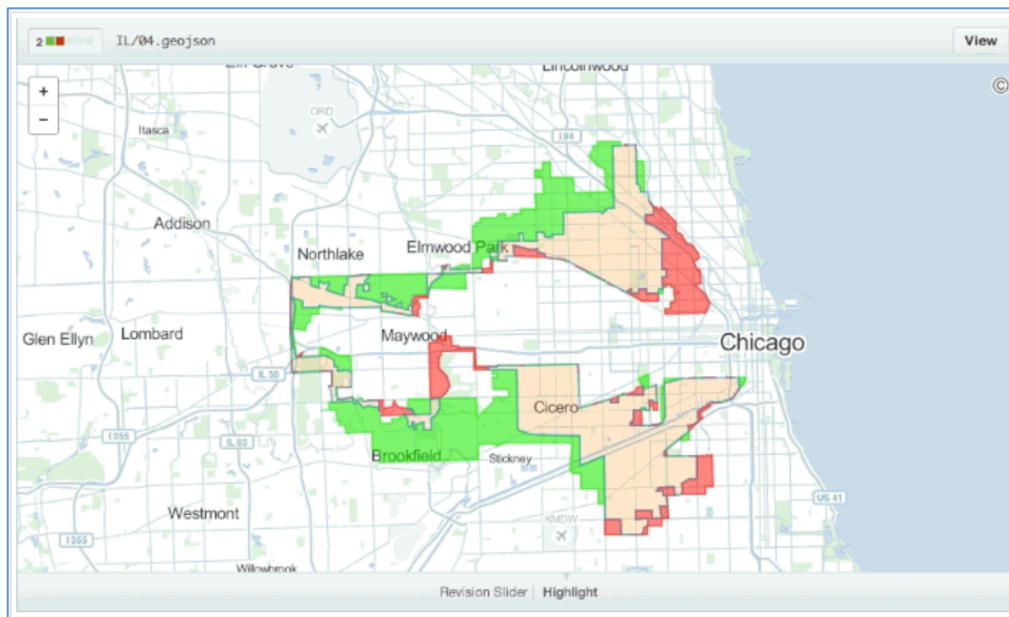


Figure 11 - Exemple de visualisation cartographique de changements territoriaux

Source : GitHub - Visualizing geospatial changes over time, <https://github.com/blog/1772-diffable-more-customizable-maps>. Illustration du différentiel entre les deux versions du territoire du 4^{ème} district de d'Illinois (congrès Américain) suite à la redéfinition de 2011 (ajout en vert, perte en rouge, maintien en saumon).

b) produire automatiquement les valeurs d'indicateurs, socio-économiques ou environnementaux par exemple, dans une nouvelle version de la partition territoriale, en utilisant un programme capable de déterminer l'opération à appliquer aux données (e.g. agrégation, désagrégation, estimation) en fonction de la nature de l'indicateur, mais aussi de la nature du changement territorial. Le fait de disposer des représentations du changement facilement exploitables par le biais de requêtes sémantiques (pour identifier toutes les unités auxquelles appliquer une même traitement car elles ont subi le même type de changement par exemple) contribue à cet objectif.

c) simuler et observer les effets d'un redécoupage territorial à des fins de gouvernance, en lien, par exemple, avec la gestion des subventions. La simulation revient ici à créer une nouvelle instance des ontologies TSN et TSN-Change (i.e. un nouveau graphe RDF⁶⁹) rendant compte du territoire redécoupé conformément au scénario étudié. Pour l'observation des effets à des fins de gouvernance, on pourra avoir recours aux traitements et aux solutions de géovisualisation respectivement évoqués en b) et a).

d) disposer de trajectoires territoriales sur le temps long, afin d'étudier la pertinence de suivre ou non un chemin similaire en termes de recomposition, extension territoriale, etc. Les descriptions du changement peuvent permettre d'identifier des patterns d'évolution des territoires, et par suite d'identifier des territoires dont les trajectoires sont similaires ou opposées pour conduire des comparaisons ou des analyses d'impact des politiques menées. Sur ce point également, on mettra à profit les possibilités d'interrogation offertes par le langage SPARQL pour exploiter la sémantique du modèle TSN/TSN-Change. Ce sera aussi l'occasion de réinvestir les techniques de raisonnement, telles que l'analyse sémantique abordée dans le cadre de la thèse de A.D. Miron [TH-2] évoquée en préambule.

⁶⁹ qui n'aura pas forcément vocation à être versé dans le Web des LOD

« Ce qui compte, ce n'est pas une œuvre,
c'est la trajectoire de l'esprit durant la totalité de la vie. »

Joan Miro (1893 - 1983)

Trajectoires de vie personnelles 3

3.1 Contexte des travaux

3.1.1 Motivations et Enjeux

La question de la représentation des trajectoires personnelles a d'abord été évoquée dans le cadre des projets ANR Biblindex⁷⁰, puis ANR SVIEME⁷¹, auxquels j'ai participé. Dans ces projets, les besoins étaient relativement simples puisqu'il s'agissait essentiellement de représenter les lieux et dates clés de la vie d'auteurs d'écrits patristiques, puis d'ouvrages médiévaux, pour en rendre compte par le biais de géovisualisations. Ces dernières devaient permettre de voir les parcours respectifs des auteurs, de détecter s'ils s'étaient par exemple croisés et/ou si des incohérences pouvaient être révélées. Les dates et lieux de naissance et de décès, ainsi que les marqueurs dans l'espace et le temps de moments importants dans la vie de ces auteurs étaient à consigner, essentiellement sous forme attributaire.

Les travaux présentés dans ce chapitre nécessitent en revanche d'élaborer une représentation plus complexe des trajectoires de vie. Une collaboration avec des chercheurs en sociologie et des urbanistes questionnant les choix résidentiels des individus est à l'origine de ces recherches. Le positionnement des motivations scientifiques de nos partenaires est présenté plus en détail dans (Villanova-Oliver *et*

⁷⁰ ANR Blanc SHS BIBLINDEIX : Analyse spatio-temporelle des références bibliques dans la littérature patristiques (2010-2014).

⁷¹ Projet ANR SVIEME (*Scientific and Visual Images in Early Modern Europe*), sept. 2016 - fév. 2018. Programme Construction de l'Espace Européen de la Recherche (MRSEI - vague 3). Un des résultats de ce projet orienté réseautage est le dépôt d'un projet H2020 nommé *SQUARE It* en mars 2018 par le consortium SVIEME (topic DT-TRANSFORMATIONS-12-2018-2020: Curation of digital assets and advanced digitisation).

al., 2019). En sociologie, l'analyse des migrations intra-urbaines est vue comme un des révélateurs de ce qui se joue dans l'histoire des individus et des groupes : leurs déplacements⁷² dans l'espace de la métropole ne sauraient se comprendre sans référence aux phases et cycles de la vie (composition/recomposition des ménages, naissances, etc.), aux modes concernant le cadre de vie, ou encore en considérant des facteurs économiques (Bonvalet and Bringé, 2012; Lelévrier, 2007). C'est dans ce cadre alliant sociologie et urbanisme qu'ont été posées les problématiques de la thèse de D. Noël [TH-5] avec l'objectif d'aller vers une connaissance plus fine des raisons qui conduisent les ménages à faire des choix de résidence, dont la succession dessine une trajectoire résidentielle. L'utilisation du terme de « *trajectoire résidentielle* » (Authier et al., 2010) ne fait pas référence qu'à l'aspect spatio-temporel de la vie résidentielle des individus. Pour les sociologues Yves Grafmeyer et Jean-Yves Authier (Grafmeyer and Authier, 2011) parler de trajectoire est plus spécifique que de décrire une simple évolution dans le temps et l'espace : « *l'utilisation du terme de "trajectoire" plutôt que celui "d'itinéraire" ou de "mobilité", revient à suggérer qu'une série donnée de positions successives n'est pas le simple fait du hasard, mais s'enchaîne au contraire selon un ordre intelligible.* »

Notre travail ambitionne de donner du sens aux processus de migration urbaine par une étude des trajectoires résidentielles, visant à comprendre les raisons sous-jacentes qui poussent les individus à changer de résidence, à quel moment de leur vie et pour quel (autre) espace. Mais les raisons d'un choix résidentiel ne peuvent être appréhendées que dans le contexte d'une approche globale des trajectoires de vie des individus, c'est-à-dire en prenant en compte certes les aspects familiaux et professionnels, mais aussi ceux relevant de la sphère des loisirs et de tous les aspects de la vie qui sont déterminants dans le choix d'un individu ou d'un ménage donné.

Les conditions du choix résidentiel ont été bien documentées dans les dernières décennies, notamment par les sociologues et les démographes. L'utilisation d'une perspective biographique, c'est-à-dire le fait d'envisager le choix résidentiel en lien avec le cours de la vie (*life course*) a grandement contribué à l'enrichissement de la modélisation de cette notion (Clark and Davies Withers, 2007). Il s'agit de considérer les choix résidentiels comme partie intégrante de la biographie des individus, pour mieux les mettre en perspective et analyser les relations entre les différents aspects de la vie des individus (Willekens, 1991). Les premiers travaux d'étude des choix résidentiels, avec une perspective biographique, s'attachent particulièrement à mettre en évidence les relations entre les différents événements qui ponctuent la vie des individus. Dans de nombreux cas, le principe est de modéliser le déséquilibre provoqué par les événements survenus dans la vie des individus (changement d'emploi, naissance,...), et de considérer que ce déséquilibre provoque un déménagement lorsqu'il atteint un certain point (Mulder and Hooimeijer, 1999). D'autres études sont consacrées plus particulièrement à l'analyse des événements qui surviennent simultanément (Mulder and Wagner, 1993).

Plus récemment, des auteurs rappellent que si la biographie des individus est souvent considérée comme une suite d'événements liés entre eux, l'importance des périodes de stabilité résidentielle ne doit pas être sous-estimée (Coulter and van Ham, 2013), car la vie des individus contient finalement relativement peu de déménagements, effectifs ou même simplement envisagés. La conclusion logique des auteurs est que, si une meilleure connaissance des choix résidentiels passe par une réponse à la question « *pourquoi les individus déménagent-ils?* », il ne faut pas sous-estimer l'importance d'une autre question : « *pourquoi ne déménagent-ils pas ?* ».

Décrire les migrations et trajectoires résidentielles peut se révéler un outil majeur de l'aménagement urbain. En filigrane, en dehors des déplacements imposés, on espère mieux comprendre pourquoi individus ou ménages quittent un endroit ou, au contraire, le rejoignent, compte tenu d'éléments de contexte personnels, professionnels, socio-économiques, culturels... La démarche vise à éliciter toute connaissance utile à une prise de décision éclairée pour, par exemple, mieux adapter l'offre de logements, ou encore favoriser le développement de services de proximité. Des données récoltées à propos de trajectoires de vie représentées de façon adéquate permettront ainsi de tester différentes hypothèses sur les causes de changement de résidences et, en regard, d'alimenter des

⁷² Il ne s'agit pas ici d'interroger les mobilités quotidiennes telles que celles des Enquêtes Ménages Déplacements mais bien d'observer des migrations dans l'espace en lien avec les choix résidentiels des populations.

décisions politiques. À titre d'exemple, dans le cadre de nos collaborations, nos collègues sociologues et urbanistes ambitionnent de confirmer que l'exode en banlieue s'explique par le manque (vécu, exprimé, quantifié) de logements adaptés en centre ville, mais aussi, pourquoi pas, par une tout autre cause non explicitée à ce jour. De tels résultats pourraient permettre la prise de décision voire l'action (réorienter l'offre de logements) en lien avec des phénomènes (exode péri-urbain) aux répercussions majeures (l'effet pollution lié aux trajets domicile-travail) (Villanova-Oliver *et al.*, 2019).

3.1.2 Modélisation des trajectoires de vie et des choix associés

Sous-jacente à la notion de migration urbaine, la notion de trajectoire fait l'objet de nombreux travaux en informatique, notamment depuis l'essor des technologies à base de localisation par GPS d'objets mobiles (Güting and Schneider, 2005) de toutes sortes (véhicules, humains, animaux, *etc.*). Au delà du stockage, dans des bases dédiées, de données GPS, les recherches se sont rapidement orientés vers l'expression de la sémantique associée à des trajectoires, à des fins d'exploration et d'analyse des mouvements observés.

La modélisation des trajectoires sémantiques (Alvares *et al.*, 2007b, 2007a) vise à enrichir les données de mobilités « brutes »⁷³, avec des informations susceptibles d'apporter du sens. La plupart des approches sémantiques consistent ainsi à annoter les différentes parties du découpage proposé pour les trajectoires spatio-temporelles. Ainsi, en s'inspirant de la « *time-geography* » (Hägerstrand, 1970), les trajectoires spatio-temporelles ont souvent été modélisé grâce à des périodes de déplacement et à des périodes d'activité, autrement qualifiées de mouvements (*move*) et d'arrêts (*stop*) (voir, par exemple, (Hu *et al.*, 2013; Parent *et al.*, 2013; Spaccapietra *et al.*, 2008; Zheni *et al.*, 2009)). Ce premier niveau de structuration est ensuite utilisé pour associer des informations comme le mode de transport lors d'un déplacement, ou la nature d'une activité lors d'une période étiquetée comme telle. Dans un travail récent portant sur la mobilité quotidienne et humaine en milieu urbain, (Jin and Claramunt, 2018) soulignent qu'en plus d'un modèle sémantique dédié à la représentation des trajectoires, des opérations et fonctions pour manipuler les données sont à proposer afin de révéler des motifs récurrents dans l'espace et dans le temps. L'approche de (Bogorny *et al.*, 2014) permet d'associer plusieurs « couches d'informations »⁷⁴ sémantiques à une même trajectoire. L'approche, formalisée en UML, repose sur un modèle qui présente une trajectoire sémantique (appelée *SemanticTrajectory*) comme une composition de sous-trajectoires sémantiques (*SemanticSubtrajectory*), chacune traduisant une vision particulière (*i.e.* un *aspect*) de la trajectoire complète. Une sous-trajectoire peut représenter la séquence (classique) de *stop* et de *move*, une autre peut définir des portions de cette trajectoire réalisées selon différents moyens de transport, et enfin, une troisième peut révéler les périodes réalisées dans des conditions météorologiques différentes. Cette approche permet donc des structurations différentes de la même trajectoire, chacune traduisant un *aspect* selon lequel elle est observée (ce que nous visons également). Cependant, il s'agit ici d'une proposition adaptée à l'enrichissement de données de trajectoires brutes, acquises pour des objets mobiles, comme en témoignent les concepts employés dans le modèle (*e.g.* une *SemanticSubtrajectory* est une composition de *SemanticPoint*, des objets estampillés temporellement, ayant une géométrie, et auxquels sont associées des mesures de vitesse et d'accélération).

Les travaux initiés par Marius Thériault font référence dans le domaine de la modélisation de trajectoire de vie. Le modèle spatio-temporel pour l'analyse des trajectoires de vie (Thériault *et al.*, 1999) présente trois trajectoires (professionnelle, familiale et résidentielle) modélisées selon une approche relationnelle, par des épisodes – des statuts stables pendant un intervalle de temps –, et des événements qui viennent altérer un ou plusieurs de ces statuts. Plus tard, le modèle sera modifié de façon à permettre de déterminer la probabilité qu'un événement survienne dans certaines conditions dans une trajectoire de vie (Thériault *et al.*, 2002). Le modèle sert également dans une étude des choix résidentiels (Vandersmissen *et al.*, 2009) avec un focus particulier sur le rôle joué par les changements d'emploi qui interviennent dans la vie professionnelle des individus. L'approche est plutôt guidée par l'aspect temporel: on s'intéresse davantage aux raisons pour lesquelles les individus vont déménager à

⁷³ Obtenues à l'aide de dispositifs dédiés habituellement sous forme d'une suite de triplets (latitude, longitude, temps) non enrichis sémantiquement.

⁷⁴ Ce terme n'est pas employé par les auteurs.

un moment donné (en fonction de leurs circonstances de vie) qu’aux raisons pour lesquelles ils quittent un lieu et en choisissent un nouveau. Ces travaux sont les plus proches des nôtres dans les problématiques qu’ils adressent.

3.1.3 Positionnement de l’approche

Dans (Ferrero *et al.*, 2016), les auteurs soulignent l’importance de disposer de représentations multidimensionnelles des trajectoires (qu’ils nomment « *multiple aspect representations* ») pour relever les défis liés à leur analyse (« *data analysis and mining* »). Ils citent les travaux de (Bogorny *et al.*, 2014) – pour la structuration, dans leur modèle, d’une trajectoire comme une composition de sous-trajectoires sémantiques – et les nôtres (Noël *et al.*, 2015), comme des exemples de contributions pour la représentation multi-aspects des trajectoires. À l’image de la plupart de travaux cités dans la section précédente et visant la sémantisation (multidimensionnelle) des trajectoires, l’approche de (Bogorny *et al.*, 2014) porte toutefois sur des trajectoires de données brutes, acquises par capteurs. Comme évoqué précédemment, ce type d’approche, même présentées par leurs auteurs comme génériques et permettant la représentation de toute trajectoire sémantique, reste trop souvent dédié à des trajectoires brutes à enrichir sémantiquement. Dans notre contexte, la notion de trajectoire diffère en ce sens qu’elle vise la représentation de la vie d’un individu, dont on souhaite représenter l’évolution sur un temps, de fait, généralement plus long que dans le cas du suivi d’objets mobiles équipés de capteurs. Parmi les différences notables que cela entraîne, nous pouvons souligner :

1. que la plupart des concepts de structuration des trajectoires (*e.g. fix* et *segments* par exemple dans (Hu *et al.*, 2013)) sont inadaptés pour notre problématique, notamment en raison des autres concepts liés qui font fortement référence à l’acquisition par capteurs (*e.g. : la Source* de type *ssn :Device*⁷⁵ associée à un *fix* dans l’ontologie de (Hu *et al.*, 2013)).
2. que la dimension géographique reste fortement exploitée comme composante majeure de la représentation des trajectoires, alors que nous avons besoin de modéliser des évolutions dans des espaces abstraits (postes occupés au cours d’une carrière, par exemple), ce qui ne requiert pas de référencement géographique. Sur ce point, nous rejoignons la position exposée dans (Spaccapietra *et al.*, 2008) selon laquelle de telles trajectoires métaphoriques (*Metaphorical trajectories*) nécessitent de repenser les relations spatiales par rapport aux trajectoires d’objets mobiles évoluant dans des espaces géographiques.
3. que les temps et les espaces à représenter dans notre contexte diffèrent de ceux traditionnellement employés. Contrairement aux données de trajectoires sémantiques d’objets mobiles, notre problématique nécessite des approches particulières de gestion de données incertaines (du fait de devoir retracer *a posteriori* le parcours d’une vie), de prise en compte de niveaux de granularités différents dans l’expression des informations mais aussi à des fins d’analyse ultérieure, de l’évolution possible (probable à l’échelle d’une vie) des espaces à référencer.

L’approche de (Thériault *et al.*, 1999) présente très clairement les fondements conceptuels sur lesquels nous nous appuyons pour la représentation des trajectoires de vie. Ces travaux ont initié une approche multidimensionnelle de la notion de trajectoire de vie et visait aussi un objectif d’exploration des choix résidentiels. Néanmoins, les propositions des auteurs n’offrent pas de support natif pour l’application de l’approche à d’autres contextes : seules trois dimensions prédéfinies sont prévues (sémantique familiale, professionnelle et résidentielle) avec, comme dimension à analyser, la dimension résidentielle. Nous visons plus de généralité dans la définition du modèle afin de nous affranchir des domaines d’application cible. Par ailleurs, concernant l’évolution à décrire dans des espaces métaphoriques, ou bien encore les enjeux liés aux représentations des temps et des espaces (respectivement points 2 et 3 ci-dessus), rien n’est explicitement considéré ou formalisé dans l’approche de (Thériault *et al.*, 1999). Enfin, proposée à une époque antérieure à l’avènement des technologies du Web Sémantique, l’approche n’a de fait pas été envisagée dans ce contexte. Les atouts des langages du Web des LOD pour la représentation de trajectoires d’objets mobiles sont désormais exploités dans de nombreuses approches (voir, par exemple, (Baglioni *et al.*, 2008; Wannous *et al.*,

⁷⁵ *ssn* pour *Semantic Sensor Network* (Compton *et al.*, 2012)

2015) pour des trajectoires d'objets mobiles) et ne peuvent aujourd'hui être ignorés dans une réponse à la problématique de représentation des trajectoires de vie.

Plus précisément, notre approche exploite les principes des patrons de conception d'ontologie (*Ontology Design Patterns*, (Blomqvist and Sandkuhl, 2005)). Ces derniers visent à décrire de manière générique une construction que l'on peut trouver de manière récurrente dans les ontologies. Par *construction*, on entend un ensemble de composants d'ontologie, organisés en concepts et prédicats, qui, une fois implémentés, constituent une partie de l'ontologie finale. Interprétant les propos de (Blomqvist and Sandkuhl, 2005), nous adoptons une démarche qui consiste à définir des *patrons dépendant d'un domaine* (celui des trajectoires de vie en l'occurrence). Il ne s'agit donc pas de patrons qui visent à répondre à un problème de modélisation considéré à un trop haut niveau d'abstraction. L'intérêt de proposer des patrons de conception d'ontologie suffisamment spécifiques est de pouvoir leur associer des outils permettant leur usage semi-automatisé. S'ils sont donc relativement spécifiques, nos patrons conservent toutefois de bonnes propriétés de généralité. D'une part, ils ont vocation à être utilisés quel que soit le domaine d'application qui nécessite la représentation et l'étude de trajectoires de vie. D'autre part, leur structuration est telle que le même patron de conception sera employé autant de fois que nécessaire pour représenter les différentes dimensions d'une trajectoire.

La thèse de D. Noël aborde ainsi la représentation des trajectoires de vie dans le Web des LOD avec l'objectif de favoriser l'opérationnalisation de l'approche mais aussi sa généralité. L'objectif d'opérationnalisation vise à répondre au cadre applicatif à l'origine de ce travail, celui de l'étude des choix résidentiels. La généralité visée s'explique de façon naturelle par la volonté systématique que nos travaux de recherche soient des solutions appropriées pour divers domaines d'applications. L'analyse de trajectoires de vie présente justement un intérêt dans différents contextes en plus de celui qui constitue le cadre applicatif de ces travaux (voir section 3.1.1). Nous avons déjà évoqué les premières pistes de réflexion amorcées dans le cadre des projets ANR Biblindex, puis SVIEME, pour le suivi de la vie d'auteurs de différentes époques. Dans le secteur médical, le suivi longitudinal des individus peut également bénéficier de connaissances relatives aux différents aspects de leur vie ; c'est par exemple ce que nous envisagerons d'intégrer, à terme, dans le cadre de la thèse de L. Toro [TH-7] qui porte sur des outils de représentations des données de santé au travail à des fins de surveillance. Connaître les trajectoires professionnelles passées des individus est pertinent pour l'analyse des pathologies recensées, afin, par exemple, de les mettre en regard des caractéristiques de carrières menées dans tel ou tel secteur (industrie agro-alimentaire, chimie, etc.).

Dans la section suivante, nous présentons les concepts clés pour la représentation des trajectoires de vie qui ont guidé la suite de nos propositions. La première contribution de ce travail (voir section 3.3) prend la forme de *patrons de conception d'ontologie* facilitant la production du modèle de représentation des connaissances adapté au cas d'étude (*i.e.* l'ontologie d'application). Cette démarche est par ailleurs outillée par un *framework* logiciel qui constitue la seconde contribution de ce travail (voir section 3.4). Il permet, outre la création de l'ontologie selon une démarche que nous décrivons, d'en assurer l'implémentation dans un *triplestore*, et d'en supporter le peuplement et l'exploitation. La section 3.5 présente une synthèse du travail réalisé et ouvre sur quelques prolongements.

3.2 Concepts clés pour la représentation des trajectoires de vie

3.2.1 Structurer des trajectoires de vie multidimensionnelles

La Figure 1 représente un extrait de la trajectoire de vie d'un individu inspiré des travaux présentés dans (Thériault et al., 1999). La partie gauche liste un ensemble (non exhaustif ici car uniquement illustratif) d'attributs⁷⁶ qui servent à caractériser la situation d'un individu : le nombre d'enfants de son foyer, le poste occupé, le type de logement, etc. La partie droite montre les valeurs prises par les différents attributs au cours du temps (on observe, par exemple, qu'après une période sans enfant, le foyer a successivement été composé d'un, puis de deux enfants). Le passage d'une valeur de l'attribut *nombre d'enfants du foyer* à une autre est marqué par un *événement* (symbolisé par une pastille étiquetée, comme E5) qui peut, ici, correspondre à une naissance. Les *événements* correspondent à quelque chose qui survient et qui peut être observé (Abler et al., 1971). Avant et après

⁷⁶ Thériault (1999) nomme les valeurs que prennent ces attributs des « statuts »

un événement (ou entre deux événements), on note une période de stabilité au cours de laquelle la valeur de l'attribut ne change pas. Dans notre exemple, on peut ainsi distinguer trois périodes dans la vie de l'individu si on s'intéresse uniquement au nombre d'enfants de son foyer. Ces périodes de stabilité sont appelées des *épisodes*. Nous considérons ainsi une trajectoire de vie comme la retranscription structurée des *épisodes* et *événements* observables dans la vie d'un individu. Des définitions formelles de ces concepts sont présentées dans (Noël *et al.*, 2016).

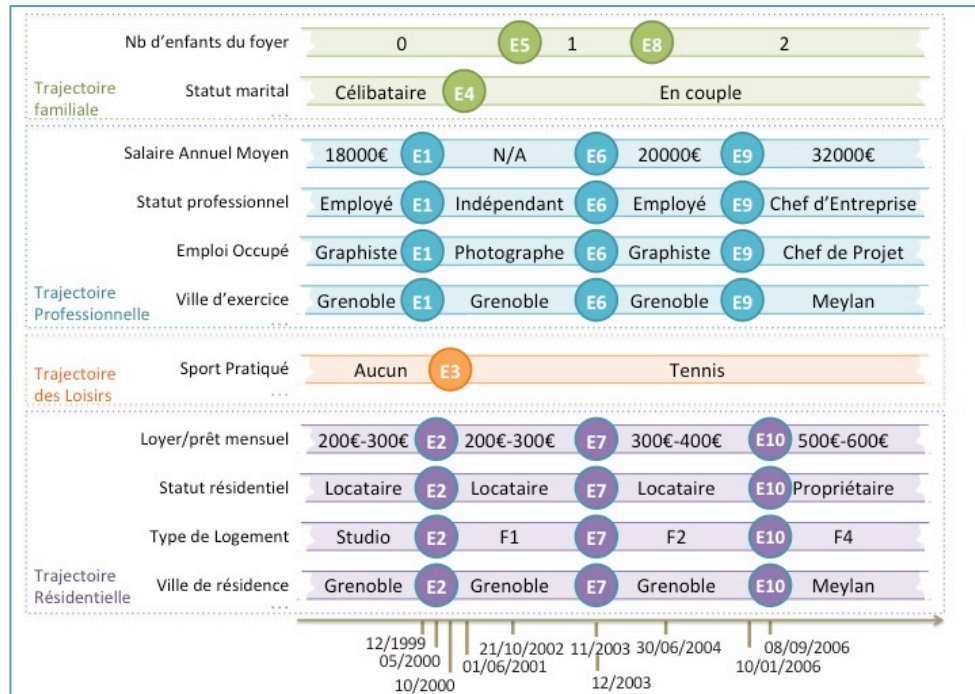


Figure 12 - Extrait d'une trajectoire de vie.

Source : Représentation inspirée de (Thériault *et al.*, 1999). Auteurs : D. Noël, M. Villanova-Oliver, 2018.

Pour définir un épisode, il est possible – mais rare – qu'un seul attribut soit pris en compte. Le cadre théorique exploité ici est celui des travaux définissant la notion de *point de vue*⁷⁷ sur une entité. Conformément à la conception la plus répandue, nous considérons que l'entité (ici, l'individu) est unique et qu'un point de vue regroupe plusieurs attributs qui contribuent à décrire le même « aspect » de la vie d'une personne (tel que l'aspect résidentiel, familial, professionnel, lié aux loisirs, *etc.*). Le regroupement d'attributs sur ce critère peut ainsi être vu comme un focus « thématique » sur la trajectoire de vie. L'union des points de vue forme une vision complète de l'entité telle que représentée pour les besoins du cas d'étude (tous les attributs appartiennent à la définition de l'identité de l'entité ou à un des points de vue).

Nous parlons de trajectoire de vie *multidimensionnelle* car chaque dimension correspond à une thématique selon laquelle on « observe » la vie de l'individu, en s'appuyant sur le regroupement d'informations qui font sens pour définir ce point de vue. Pour revenir à l'exemple, la trajectoire familiale correspond au regroupement des attributs *Nb d'enfants du foyer* et *Statut marital* ; la trajectoire résidentielle est représentée par le regroupement des attributs *Loyer/Prêt mensuel*, *Statut Résidentiel*, *Type de Logement*, *Ville de Résidence* (cf. Figure 1).

Assez naturellement, les événements et les épisodes peuvent être également qualifiés de *thématiques* parce qu'ils sont en lien avec des attributs eux-mêmes attachés à une thématique. Par exemple, E2 dans la Figure 12, traduit un événement dit *résidentiel* car il impacte des attributs de la trajectoire résidentielle. Un nouvel épisode thématique, défini à partir de tous les attributs de la trajectoire associée, apparaît dès lors qu'un au moins des attributs de celle-ci a subi un changement de valeur résultant d'un événement. Ainsi, pour la trajectoire familiale illustrée dans la Figure 1, on observe 4 épisodes :

⁷⁷ Points de vue et trajectoires thématiques sont à comprendre comme des synonymes dans ce travail.

- Un épisode avant E4 : l'individu est célibataire, sans enfant au foyer ;
- Un épisode entre E4 et E5 : l'individu est en couple, sans enfant au foyer ;
- Un épisode entre E5 et E8 : l'individu est en couple, avec 1 enfant au foyer ;
- Un épisode depuis E8 : l'individu est en couple, avec 2 enfants au foyer ;

Un épisode correspond à l'état stable d'une personne observée d'un point de vue thématique pendant un intervalle de temps. Ce choix de modélisation impose de bien choisir l'information à représenter (quels sont les attributs pertinents), mais aussi de déterminer le degré de précision attendu puisque tout changement de valeur se traduit par la création d'un nouvel épisode. La section suivante discute cette question de la granularité des données.

3.2.2 La granularité de l'information considérée

Tous les épisodes et les événements, et par suite les trajectoires, sont intrinsèquement temporels car liés à une ligne de temps selon laquelle on peut les ordonner. La datation des événements permet de borner les épisodes des trajectoires. De façon naturelle, le temps va donc jouer un rôle de référentiel commun à toutes les trajectoires thématiques qui permettra l'exploitation croisée des informations de ces dernières. L'espace peut lui aussi jouer un rôle de référentiel commun exploité dans la représentation et l'étude des trajectoires. Néanmoins, la dimension spatiale (au sens géographique du terme) n'est pas intrinsèque à toute information comme l'est le temps : si toutes les informations sont valides sur une période de temps, elles n'ont en revanche pas toutes un caractère spatial évident. Il est, bien sûr, des trajectoires thématiques qualifiables de spatiales au sens où elles intègrent des attributs dont les valeurs sont des informations de nature géographique (par exemple, la localisation des logements successifs). Mais, dans d'autres cas, le terme de *trajectoire* est utilisé comme une métaphore : pour caractériser l'évolution dans un *espace abstrait* (par exemple, les différents métiers occupés au cours d'une carrière peuvent s'exprimer dans l'espace abstrait décrits par la nomenclature des CSP⁷⁸ de l'INSEE). Notre modèle permet ainsi de caractériser de manière générique toute évolution, qu'elle s'inscrive dans un espace abstrait ou géographique.

Les niveaux de granularité auxquels sont représentées les informations déterminent les phénomènes qui seront révélés. Dans la lignée de travaux tels que ceux de (Hornsby and Egenhofer, 2002), on ne peut ignorer l'importance des niveaux de granularité temporelle pour la représentation d'objets mobiles. Un premier enjeu est de définir la granularité temporelle adaptée à la modélisation de la trajectoire de vie. Pour (Jensen et al., 1998), un *chronon* est une durée de temps minimum, la plus petite unité temporelle considérée pour marquer temporellement un événement. Dans le cas de la trajectoire de vie, un chronon pertinent peut être le jour, cette unité permettant de situer des événements marquants avec une précision qui fait sens pour un individu (comme la date d'un mariage, de la naissance d'un enfant, *etc.*). À l'échelle d'une vie, on pourra retenir une granularité moins fine, telle que le mois ou l'année, pour situer un événement pour lequel la précision en termes de jour est impossible (oubli de la date précise), ou non forcément requise (par exemple, pour situer quand a débuté le projet de déménager). Le modèle doit donc supporter des degrés de précision différents dans la datation et, en ce sens, le recours à l'ontologie OWL-Time⁷⁹ sera utile du fait de la flexibilité qu'elle permet.

Un raisonnement similaire est appliqué pour déterminer le niveau de granularité auquel il convient de considérer l'information de nature géographique. Néanmoins, contrairement au temps, l'espace est un référentiel qui peut évoluer au cours du temps. On ne redéfinit pas ce qu'est un mois par exemple, alors qu'une commune peut évoluer (fusionner avec une autre, se scinder en deux, *etc.*) ce qui oblige à gérer des évolutions territoriales sur un temps long (*i.e.* à l'échelle d'une vie en l'occurrence). Ici, référencer les territoires mentionnés dans une trajectoire de vie selon l'approche présentée dans le chapitre 2 (thèse de C. Bernard [TH-6]) est pertinent. Parallèlement, l'impact du découpage spatial retenu pour rendre compte des données, connu sous le nom du *Modifiable Area Unit Problem* (MAUP, (Openshaw and Taylor, 1979)), est un argument qui nous amène à préconiser le recueil d'une information spatiale aussi fine que possible, offrant par suite plus de latitude dans le choix des niveaux

⁷⁸ Nomenclature de catégories socioprofessionnelles / CSP (<https://www.insee.fr/fr/metadonnees/definition/c1758>)

⁷⁹ <https://www.w3.org/TR/owl-time/>

auxquels seront réalisées des agrégations spatiales dans l'étude des trajectoires de vie. La définition selon l'approche TSN et TSN-Change de nomenclatures visant à décrire des territoires selon des découpages fins (de niveau communal comme dans la LAU, voire à l'échelle des métropoles en distinguant les quartiers), est envisagée dans le cadre de ce travail.

La recherche d'une granularité fine de description des informations est un principe également valable pour tous les attributs qui peuvent être adossés à des catégories structurées de façon hiérarchique (comme les CSP de l'INSEE). Dans ce sens, le recours à des vocabulaire, contrôlés sera privilégié autant que possible.

La question de la granularité de l'information dans la représentation des trajectoires de vie reste toutefois à questionner au moment de la définition du modèle (voir section 3.4.1), en regard des objectifs et besoins de l'étude. La recherche de granularité la plus fine n'est pas forcément souhaitable pour toutes les informations, du fait de son impact sur la stabilité des épisodes. Ainsi, pour revenir à notre exemple, ce n'est pas le montant exact d'un loyer qu'il s'agira de stocker, mais plutôt l'intervalle de valeurs d'appartenance de loyer ou toute autre catégorisation qualitative de l'information (loyer modéré, élevé, *etc.*) qui présentent l'avantage d'accroître la stabilité des épisodes.

Les événements et épisodes, dans notre approche, servent à structurer une trajectoire thématique et plusieurs trajectoires thématiques composent une trajectoire de vie. Notre approche dépasse toutefois le cadre d'une relation partie-tout, par le rôle de *facteurs explicatifs* donné aux événements et aux épisodes dans la représentation d'une trajectoire de vie.

3.2.3 Lier l'approche multidimensionnelle et les facteurs explicatifs

Pour étudier la trajectoire résidentielle d'un individu (*i.e.* identifier et comprendre la succession de ses choix résidentiels dans le temps et dans l'espace), il convient de la considérer comme une partie (*i.e.* une trajectoire thématique) intégrée dans une vision plus large (*i.e.* la trajectoire de vie de l'individu). Dans la représentation proposée Figure 12, il apparaît clairement que les différentes trajectoires thématiques coexistent dans cette vision multidimensionnelle, étant reportées sur une même ligne de temps. Dès lors, le temps étant un référentiel commun 'naturel', l'utiliser comme filtre dans l'étude des trajectoires est possible pour observer, par exemple, la concomitance d'événements, ou toute autre relation liant événements et/ou épisodes (en exploitant notamment l'algèbre de Allen (Allen, 1984) dans des requêtes). Nous pouvons également exploiter de façon analogue le caractère spatial d'informations qui seraient représentées dans certaines trajectoires thématiques.

Notre approche dépasse cette possibilité et vise à établir des connexions entre épisodes et événements attachés à des thématiques différentes, dès lors que certains jouent le rôle de facteurs explicatifs d'autres événements survenus dans la trajectoire de vie d'un individu. Cela permettra une compréhension plus fine des trajectoires personnelles. Par exemple, deux événements – un déménagement et un changement de travail – peuvent apparaître au même moment dans la vie d'un individu. Si une requête permet de retrouver leur concomitance, elle ne permet pas de l'expliquer si le seul modèle disponible est celui de la juxtaposition des trajectoires thématiques. Une illustration simple est la suivante : un déménagement (point de vue résidentiel) peut-être expliqué par une mutation professionnelle (point de vue professionnel). Mais l'inverse peut tout aussi bien retranscrire une autre réalité, elle aussi plausible : le changement dans la carrière professionnelle est le résultat d'un déménagement. Au-delà de cet exemple trivial, un enjeu réside dans la représentation de facteurs explicatifs de natures diverses, éventuellement combinés, intervenant à différents degrés et en lien avec n'importe quelle thématique.

Deux principes régissent notre approche à base de *facteurs explicatifs*. Le premier principe est qu'un facteur explicatif vient expliquer *un autre événement* de la trajectoire, non forcément attaché au même point de vue. En cela, nous rejoignons les conceptions sociologiques évoquées en section 3.1.1 (p37). Nous considérons qu'un événement peut être expliqué soit :

- par d'autres événements de vie : par exemple, l'individu Bruno déménage (*événement expliqué*) car il a obtenu une promotion (*événement qui est un facteur explicatif*). Cet exemple montre comment un lien est établi entre la trajectoire résidentielle et la trajectoire professionnelle.

- par les caractéristiques d'un épisode : par exemple, Bruno déménagement (*événement expliqué*) car son précédent appartement était caractérisé par une surface trop petite (*épisode qui est un facteur explicatif, ici du fait de la valeur de l'attribut descriptif surface*).
- par des combinaisons d'événements et/ou d'épisodes : par exemple Bruno déménagement (*événement expliqué*) car i) il a obtenu une promotion (*événement qui est un facteur explicatif*), ii) il a eu un enfant (*autre événement qui est un facteur explicatif*) et iii) son appartement précédent était trop petit (*épisode qui est un facteur explicatif* notamment du fait d'un nombre de chambres insuffisant, traduit par un attribut nbChambres=1).

Des liens de causalité entre événements et épisodes attachés à différents points de vue d'une même trajectoire sont ainsi établis par le biais de facteurs explicatifs. L'expression de ce type de relation permet donc l'analyse complémentaire des trajectoires de vie que nous avons évoquée en début de section. Si nous pouvons, bien sûr, explorer les trajectoires en observant, par exemple, des successions temporelles (l'individu a habité ici, puis là) ou de concomitances spatiales (l'individu travaillait et résidait dans le même quartier), notre modèle permet également d'aller vers une certaine compréhension des raisons de ces observations.

Le deuxième principe de l'approche est le suivant : ces raisons sont recueillies directement auprès des individus⁸⁰ sous forme de facteurs explicatifs, permettant ainsi des *analyses des trajectoires de vie éclairées par les individus eux-mêmes*. Ainsi, pour une compréhension plus fine, les facteurs explicatifs s'accompagnent d'une explication textuelle fournie par l'individu et d'une pondération permettant de relativiser le poids de différents facteurs dans la survenue d'un événement expliqué.

Les facteurs explicatifs peuvent être internes ou externes, c'est-à-dire qu'ils peuvent soit être relatifs aux circonstances de vie de l'individu (comme illustrés précédemment avec le cas de l'individu Bruno) soit dépendre de circonstances qui n'y sont pas liées directement. Les facteurs explicatifs de la trajectoire de vie peuvent, en effet, également être trouvés dans ce qui relève de la qualité de l'environnement (économique, social, *etc.*) dans lequel évolue l'individu. Ces facteurs explicatifs externes ne sont représentés que dans la mesure où l'individu se sent impacté par ceux-ci et l'exprime comme une (des) raison(s) ayant entraîné un événement dans sa trajectoire. En toute généralité, le fait que des conditions de prêts bancaires soient favorables peut être déterminant pour un achat immobilier ou non. Si de tels facteurs externes peuvent être envisagés comme explications plausibles d'un événement, c'est bien le fait qu'ils soient explicitement mentionnés par l'individu comme facteur explicatif à un moment donné qui nous intéresse ici.

Dans les sections suivantes, nous présentons les contributions de la thèse de D. Noël qui s'appuient sur une approche ontologique ouverte sur le Web des LOD, afin d'organiser les différents éléments que nous venons de mettre en exergue.

3.3 Modélisation de trajectoires de vie : l'approche LTOP

L'approche que nous proposons, nommée *Life Trajectory Ontology Patterns* (LTOP), repose sur des patrons de conception génériques que nous avons définis, ainsi qu'une démarche pour l'application de ces patrons. Les descriptions des patrons de conception d'ontologies sont accessibles à <http://lig-tdcge.imag.fr/ltop>. Dans la suite, le préfixe *ltop:* utilisé dans les figures indique que les concepts associés sont ceux de notre approche. Dans le texte nous l'omettons pour plus de lisibilité.

Le premier patron décrit les concepts de *trajectoire thématique*, d'*épisode* et d'*événement* essentiellement, ainsi que les relations qui les lient. Le second patron sert à caractériser le concept de *facteur explicatif* en établissant des liens *ad hoc* avec les concepts d'épisode et d'événement décrits à l'aide du premier patron. Il permet de doter tout événement associé à une trajectoire thématique de la capacité à être utilisé comme facteur explicatif dans l'ontologie.

⁸⁰ Cela peut légitimement appeler à des débats (caractère possiblement biaisé de l'introspection, dissonance cognitive) que nous n'aborderons néanmoins pas dans ce manuscrit.

3.3.1 Le patron de création d'une trajectoire thématique

3.3.1.1 Cas général

Le design pattern d'ontologies *Thematic Trajectory Pattern* est présenté dans la Figure 13, dans l'encadré en pointillés. Ce patron se raccroche à un autre patron (partie supérieure de la figure) visant :

- à traduire la relation de composition entre une trajectoire de vie (*LifeTrajectory*) et des trajectoires thématiques (*ThematicTrajectory*) par le prédicat *hasThematicTrajectory*.
- à représenter la relation d'appartenance entre une *ThematicTrajectory* et un individu qui pourra être représenté comme une instance de la classe *foaf:Person*⁸¹ par une propriété *hasOwner*. Cette possibilité est donnée mais reste optionnelle dans la mesure où des conditions d'anonymat peuvent être imposées.

Les concepts principaux qui composent le *design pattern* sont : les épisodes (*Episode*) et les événements (*Event*) dont est constituée une *ThematicTrajectory* grâce aux propriétés *hasEpisode* and *hasEvent*.

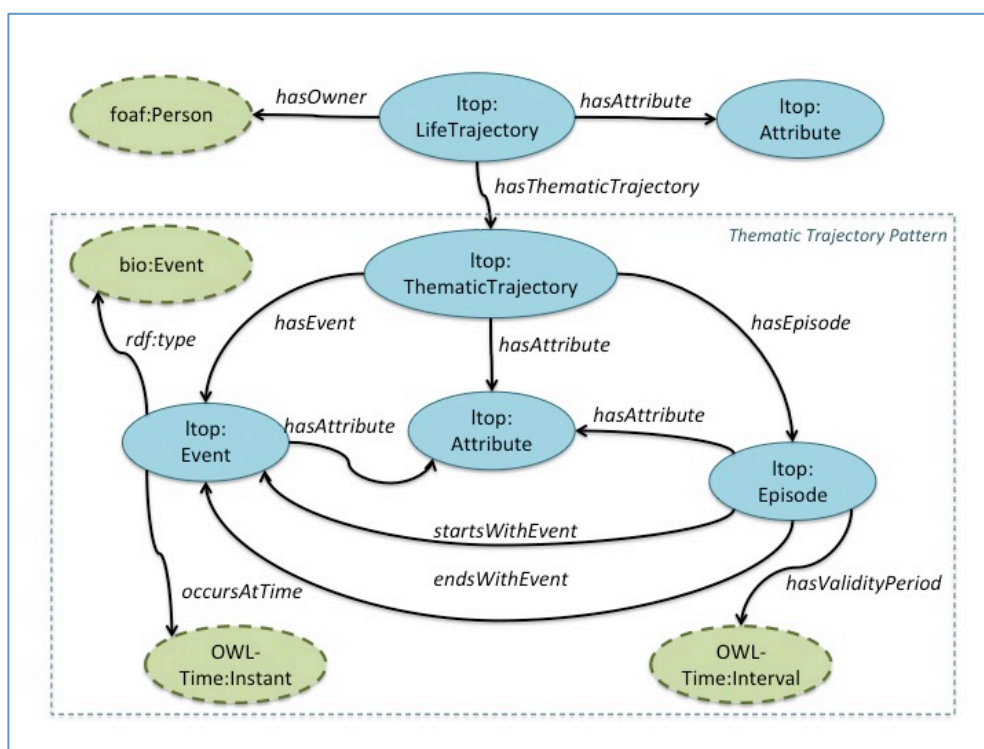


Figure 13 - Représentation simplifiée du *Thematic Trajectory Pattern*

Source : d'après (Noel et al., 2017). Auteur : M. Villanova-Oliver, 2018.

Un événement *Event* de notre approche peut-être de type *bio:Event*. Le préfixe *bio* fait référence au vocabulaire *BIO*⁸² proposé pour la description d'information biographique. S'il recense des événements de la sphère personnelle (comme une naissance, un mariage, etc.) ou encore professionnelle (promotion, perte d'emploi, etc.), ce vocabulaire n'aborde pas toutes les thématiques envisageable (dont résidentielle par exemple). Dès lors, nous intégrons la possibilité de définir de nouveaux types d'évènement par le biais de la classe *Event* (en donnant ses attributs – voir plus loin – dont une date d'occurrence). Selon les besoins de l'application finale, des ontologies génériques conceptualisant les événements peuvent être exploitées pour spécialiser notre classe *Event* (voir (Troncy et al., 2010) pour une revue).

⁸¹ L'ontologie Friend of a Friend permet de décrire les individus et relations entre eux, notamment dans le cadre des réseaux sociaux (<http://xmlns.com/foaf/spec/>). Cet aspect relationnel n'a pas encore été exploité dans nos travaux (d'un point de vue opérationnel), mais ce choix de représentation permet d'envisager l'augmentation des possibilités d'analyse ultérieures de trajectoires, en tenant compte du réseau social d'un individu dans des requêtes SPARQL.

⁸² <http://vocab.org/bio/>

Un *Event* marque le début d'un épisode (*startsWithEvent*), mais aussi la fin d'un épisode (*EndsWithEpisode*). L'événement est daté grâce au prédicat *occursAtTime* dont l'objet (*range*) est un *OWL-Time:Instant* qui pourra être décrit au niveau de granularité souhaité. L'épisode possède une période de validité (*hasValidityPeriod*) représentée par un type *OWL-Time:Intervalle* offrant la même souplesse.

Dans ce patron, apparaissent également la classe *Attribute* et le prédicat *hasAttribute* ayant comme sujet les concepts *LifeTrajectory*, *ThematicTrajectory*, *Episode* et *Event*. *Attribute* et *hasAttribute* sont définis de manière générique pour représenter le fait que les concepts, sujets du prédicat, doivent être associés aux propriétés qui font sens pour la thématique décrite (comme par exemple les attributs illustrant les différentes trajectoires dans l'exemple introductif, voir Figure 12). Les utilisateurs de ce patron pourront définir les sous-classes et prédicats associés en fonction des besoins. Cette pratique, notamment adoptée par (Hu *et al.*, 2013), garantit la flexibilité nécessaire en vue de la réutilisation du patron.

3.3.1.2 Trajectoire thématique à composante géographique ou métaphorique

Si la trajectoire thématique à modéliser possède une composante géographique, une extension du patron illustrée par la Figure 14 est proposée (cadre en pointillés, à gauche). Une trajectoire à composante géographique *GeoTrajectory* est une spécialisation de la classe *ThematicTrajectory*. La particularité de *GeoTrajectory* est portée par l'ajout d'une classe *GeoEpisode*, spécialisation de la classe *Episode*, pour laquelle on impose la relation (prédicat *hasGeoFeature*) avec un objet de type *Geo:Feature* défini par l'ontologie GeoSPARQL. C'est notamment par le biais de cette classe que nous pouvons envisager d'exploiter ici les travaux décrits dans le chapitre 2 (les unités territoriales étant des *Geo:Feature*).

Selon un principe similaire (voir Figure 14, cadre en pointillés, à droite), le support pour la représentation de trajectoires métaphoriques est assuré par le recours à une classe *MetaphoricTrajectory*, spécialisation de la classe *ThematicTrajectory*. Une classe *MetaphoricEpisode*, spécialisation de la classe *Episode*, est définie et reliée par un prédicat *hasSKOScollection* avec un objet de type *skos:Collection* tel que proposé dans l'ontologie SKOS⁸³. Ce choix est motivé par le fait que l'étude de l'évolution d'une position dans un espace abstrait requiert que cet espace soit structuré, soit sous la forme de listes ordonnées de valeurs traduisant, par exemple, une progression, soit en une nomenclature hiérarchique.

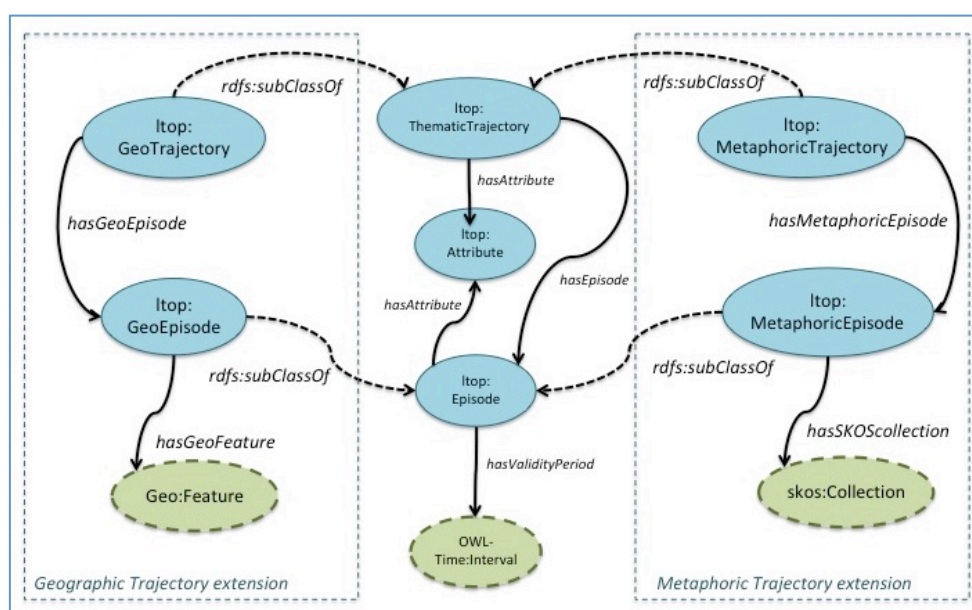


Figure 14 – Extensions pour les trajectoires à composante géographique et métaphorique

⁸³ <https://www.w3.org/TR/skos-reference/#collections>

À ce stade, nous avons montré comment sont structurées les trajectoires thématiques, géographiques et métaphoriques, qui participent à la représentation d'une trajectoire de vie selon notre approche. Nous présentons à présent le patron de conception dédié à l'expression des facteurs explicatifs.

3.3.2 Le patron d'expression de facteurs explicatifs

3.3.2.1 Cas général

Le *design pattern* d'ontologie *Trajectory Explanatory Factors Pattern* permet de modéliser des facteurs explicatifs qui seront utiles pour l'interprétation des trajectoires. Nous rappelons que dans notre approche, les facteurs explicatifs reposent sur deux principes (voir section 3.2.3) :

- ils servent expliquer des événements de la trajectoire ;
- ils sont exprimés par les individus eux-mêmes.

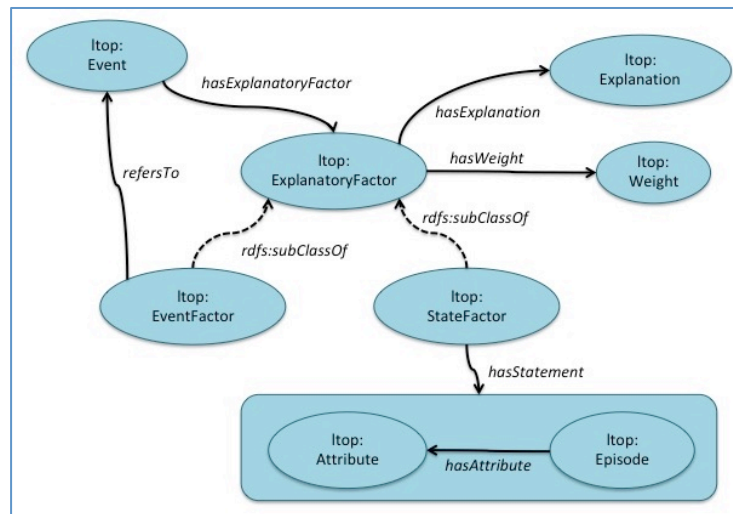


Figure 15 - Représentation schématique du *Trajectory Explanatory Factors Pattern*

Source : d'après (Noel et al., 2017).

La Figure 15 donne une représentation schématique de ce patron. La classe *Event*, en tant qu'événement d'une trajectoire thématique à expliquer, est liée au concept *ExplanatoryFactor* par le prédicat *hasExplanatoryFactor*. Ce concept est associé à une explication fournie par l'individu (*Explanation*) et un poids (*Weight*) qui permet de relativiser la part de plusieurs *ExplanatoryFactor* qui seraient associés à un même événement. Lors de l'application du patron, l'utilisateur décide de la façon dont devrait être fournies ces informations (e.g. texte libre ou vocabulaire contrôlé en fonction du type de l'événement pour le concept *Explanation*, pondération entre 0 et 1 ou échelle qualitative pour le concept *Weight*).

3.3.2.2 Distinction entre *EventFactor* et *StateFactor*

Le concept *ExplanatoryFactor* se spécialise en deux sous-concepts qu'il est possible de combiner (voir p44 et suivante) :

- *EventFactor*, pour l'explication d'un événement par un autre événement de vie comme dans l'exemple « Bruno déménagement (événement expliqué) car il a obtenu une promotion (événement qui est le facteur explicatif) ». Dans ce cas, le concept *EventFactor* fait référence au facteur explicatif (un *Event*⁸⁴) par le prédicat *refersTo*. Une illustration simplifiée est donnée par la Figure 16 (voir la légende de la figure).

⁸⁴ Un ensemble d'axiomes définit les règles de construction qui s'appliquent dans les patrons. Notamment ici, si *hasExplanatoryFactor(Evt1, EvtFactor_A)* et *refersTo(EvtFactor_A, Evt2)* sont observés alors *Evt1* et *Evt2* sont deux instances différentes de *Event*.

- *StateFactor*, pour l'explication d'un événement par les caractéristiques d'un épisode comme dans l'exemple « Bruno déménage (événement expliqué) car son précédent appartement a une surface trop petite (épisode qui est un facteur explicatif, ici du fait de la valeur de l'attribut descriptif surface) ». Une illustration simplifiée est donnée par la Figure 17 (voir la légende de la figure). Cette propriété utilise la réification *RDF*, qui permet de faire une déclaration à propos d'une autre déclaration (un triplet sujet-prédicat-objet se voit assigner un URI et est traité comme une ressource que l'on peut donc utiliser dans une autre déclaration).

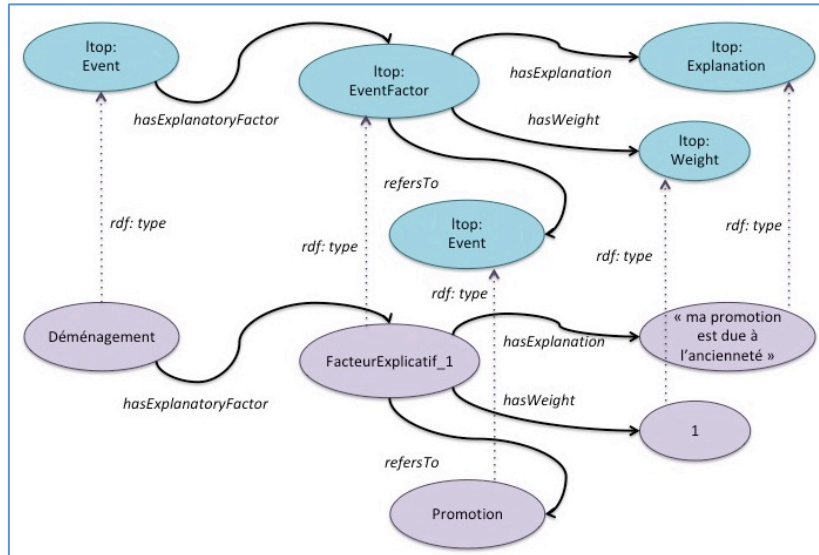


Figure 16 - Illustration simplifiée d'un facteur explicatif de type *EventFactor*.

Du fait de la simplification de la figure, il n'apparaît pas que les événements Déménagement et Promotion appartiennent à des trajectoires thématiques différentes (respectivement Résidentielle et Professionnelle). De même, des libertés sont prises ici avec la représentation des valeurs pour les concepts Explanation et Weight. Pour ce dernier, on considère que la valeur 1 traduit une explication totale du déménagement par la promotion.

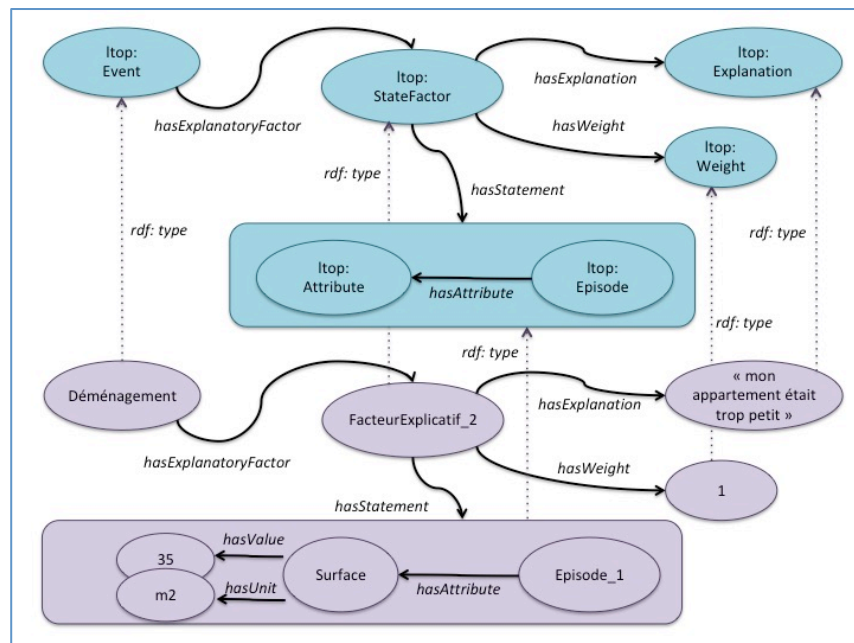


Figure 17 - Illustration simplifiée d'un facteur explicatif de type *StateFactor*.

Du fait de la simplification de la figure, il n'apparaît pas que l'événement Déménagement et l'Episode_1 appartiennent à la trajectoire thématique Résidentielle. De même, des libertés sont prises ici avec la représentation des concepts Weight (même interprétation que précédemment) et Surface.

3.3.2.3 Facteurs explicatifs externes à la trajectoire personnelle

Dans les propositions précédentes, le concept *ExplanatoryFactor* s'appuie sur des faits (événements ou caractéristiques d'épisodes) appartenant à la trajectoire de l'individu. Notre approche supporte également l'expression de facteurs externes à cette trajectoire personnelle, et ce, de deux façons.

D'une part, l'expression de facteurs explicatifs issus des trajectoires d'autres individus, décrites selon l'approche LTOP est possible. Pour cela, le concept *EventFactor* se spécialise en *InternalEventFactor* et *NetworkEventFactor*. Un *InternalEventFactor* est utilisé lorsque l'événement « expliquant » appartient à la trajectoire du même individu. Dans l'exemple précédent, la promotion de Bruno est donc en toute rigueur à décrire par ce concept *InternalEventFactor*. Un *NetworkEventFactor* est associé à un événement explicatif appartenant à la trajectoire d'un autre individu. Ainsi, si ce n'est pas la promotion de Bruno mais celle de sa compagne qui est à l'origine du déménagement expliqué dans la Figure 16, alors la classe *NetworkEventFactor* est à privilégier pour référencer un événement hors de la trajectoire de l'individu, en combinaison avec l'exploitation des liens entre individus décrits par le biais de l'ontologie *foaf* notamment. Si cela permet d'augmenter le pouvoir explicatif du modèle, cela nécessite néanmoins des dispositions particulières en cas d'anonymisation souhaitée des trajectoires de vie.

D'autre part, l'expression de facteurs explicatifs doit pouvoir s'appuyer sur l'évocation d'événements extérieurs à toute trajectoire définie selon notre approche. En cela, l'ouverture sur le Web des LOD est exploitée. Dans ce but, le concept *EventFactor* est également spécialisé en un concept *ExternalEvent* qui peut-être une instance de tout événement décrit par des ontologies telles que LODÉ⁸⁵ (Troncy et al., 2010) ou encore The Event Ontology⁸⁶, dédiées à la description d'événements historiques ou de type manifestation culturelle, sportive, ou autre.

Nous avons présenté des patrons de conceptions d'ontologie qui permettent la construction d'une ontologie dédiée à la représentation de trajectoires de vie multidimensionnelles et permettant d'en expliquer les différentes étapes. La section suivante illustre comment cette approche est opérationnalisée par la proposition d'un *framework* capable d'exploiter ces patrons et les ontologies produites.

3.4 Framework dédié à la gestion de trajectoires de vie

Nous proposons une architecture logicielle (voir Figure 18) qui tire ainsi profit du Web Sémantique tout au long du processus présenté ci-dessous. Cette architecture est composée de quatre modules offrant des fonctionnalités qui peuvent être invoquées via des API (*Application Programming Interface*). Ces fonctionnalités sont le support d'une démarche qui comprend quatre étapes illustrées par les pastilles numérotées de 1 à 4 dans la Figure 18: ① modélisation de la trajectoire par un expert du domaine d'application cible en utilisant le module *Ontology Manager*; ② acquisition des données qui viendront alimenter le modèle par le biais du module *Collect Manager*; ③ enrichissement des informations collectées grâce au *Triple Store Manager* et ④ exploitation des données par des experts à des fins d'analyse à travers le module *Exploitation Manager*. Nous présentons chacune de ces étapes ci-après, en détaillant plus particulièrement les deux premières qui ont fait l'objet des développements réalisés à ce jour.

⁸⁵ <http://linkedevents.org/ontology>

⁸⁶ <http://motools.sf.net/event/event.html>

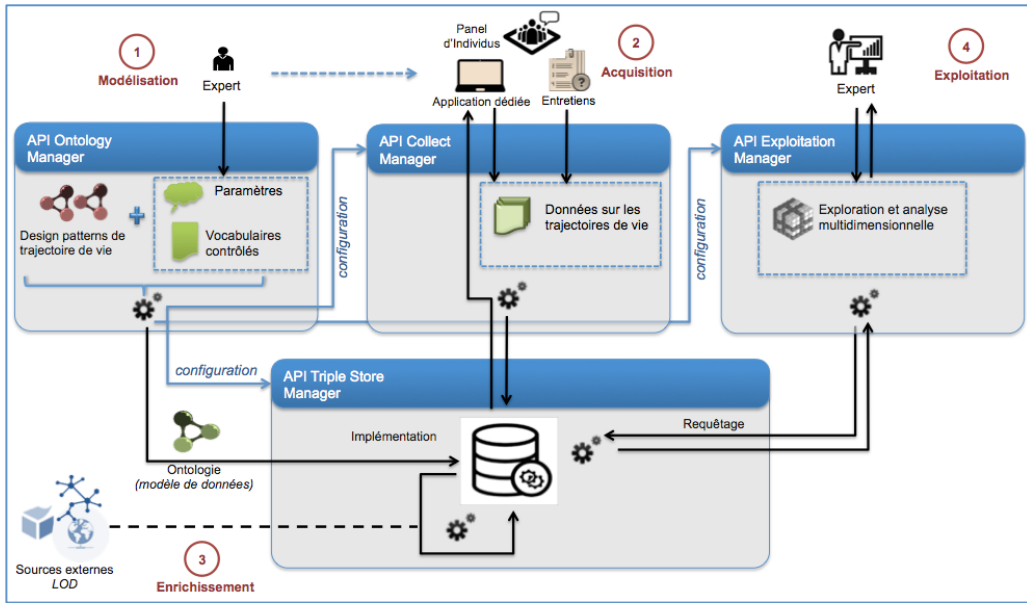


Figure 18 - Architecture pour la représentation, la collecte et l'analyse de trajectoires de vie

Source : M. Villanova-Oliver et D. Noël, 2018

3.4.1 Création du modèle d'ontologie d'application

Le modèle d'une trajectoire de vie dépend *in fine* fortement du domaine d'application cible. Le premier module *Ontology Manager* de notre architecture (voir Figure 18) permet à un expert de modéliser facilement une trajectoire de vie multidimensionnelle en rapport avec son domaine d'expertise.

Lors de la phase de création du modèle, un expert souhaitant modéliser une trajectoire de vie pour un domaine d'application particulier utilise le module appelé *Ontology Manager*. Ce composant logiciel offre des procédures automatisées (Noël et al., 2017) assurant, à partir des patrons, la production de l'ontologie d'application servant de support à l'acquisition, l'enrichissement et l'exploitation des données. La Figure 19 illustre le processus de modélisation qui nécessite que l'expert fournisse les paramètres nécessaires à la création de l'ontologie.

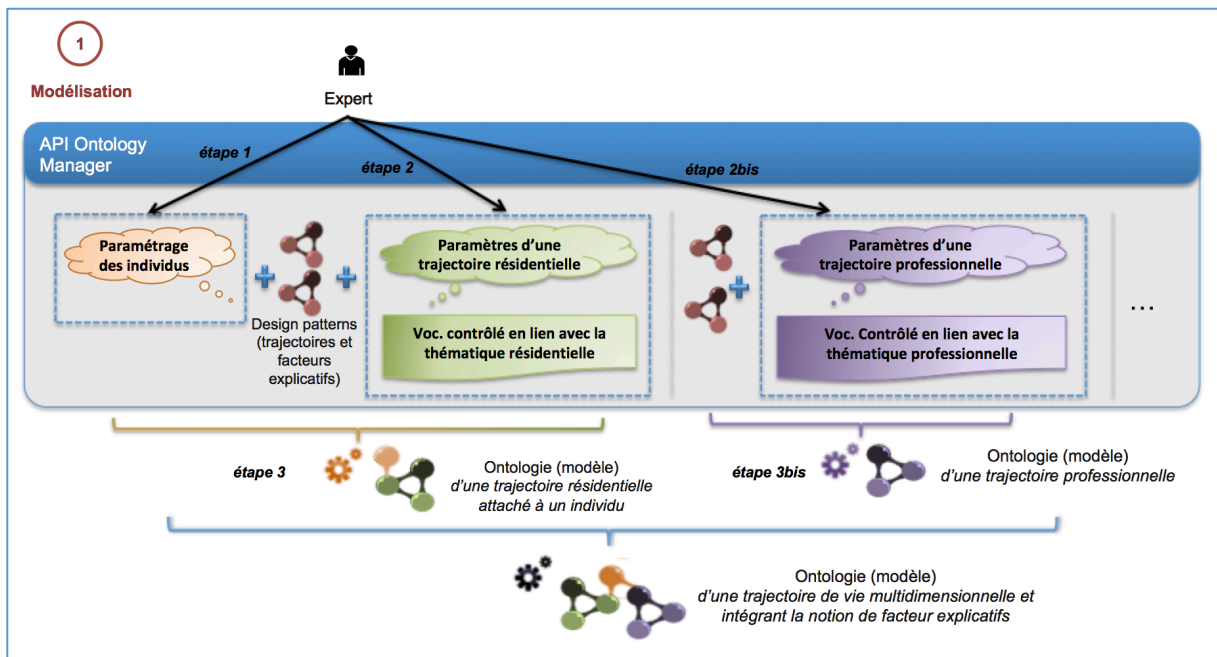


Figure 19 - Représentation de étapes de la démarche de modélisation

Le paramétrage des attributs permettant de représenter le « propriétaire » d'une trajectoire de vie est d'abord amorcé (étape 1). Il s'agit de fournir les éléments qui serviront à décrire les individus pour lesquels on crée l'ontologie : à minima un code, invariant au cours du temps et non traçable si le respect de l'anonymat est souhaitable, sera utilisé. D'autres attributs, permettant des analyses socio-démographiques, par exemple, peuvent être fournis mais il doit s'agir d'informations invariantes (une année de naissance par exemple) et non d'attributs sujets à évolutions qui seront définis dans les trajectoires thématiques appropriées.

La première trajectoire thématique est ensuite paramétrée (étape 2). L'expert donne : un nom (utilisé pour générer les classes de l'ontologie de trajectoire pour la thématique créée, par exemple *résidentielle* dans la Figure 19), son type (trajectoire thématique simple, géographique, ou métaphorique), la liste des différents attributs descriptifs qui participeront de la définition des épisodes de cette thématique. Pour chaque attribut, l'expert fournit un nom et son type. Pour ce dernier, on pourra recourir à un vocabulaire contrôlé. Ce sera notamment le cas pour les attributs géographiques (pour désigner une ville de résidence par exemple, on utilisera des ontologies de référence), et ainsi on pourra bénéficier implicitement et automatiquement de la connaissance de la structuration administrative hiérarchique d'organisation du territoire lors des analyses. Le paramétrage consiste également à fournir les éléments de description des événements. En particulier, des typologies d'événements pertinents pour une thématique peuvent être fournies (par exemple, pour la thématique résidentielle, on aura certainement à considérer des *déménagements*, des *achats de biens*, des *travaux*, etc.).

À partir de ces paramètres et du patron de conception, une première partie de l'ontologie de trajectoire de vie est donc automatiquement créée grâce à des programmes dédiés (étape 3). Les vocabulaires contrôlés sont stockés également de manière automatique en vue d'une utilisation ultérieure lors des phases de collecte et d'exploitation des données.

Les étapes 2bis et 3bis symbolisent le fait que l'expert va pouvoir réappliquer le processus de paramétrage et de génération de l'ontologie autant de fois que nécessaire (*i.e.* autant qu'il y a de trajectoires thématiques à créer pour les besoins de l'étude). Si la première étape a mis la trajectoire résidentielle au premier plan de la modélisation – car, pour notre contexte applicatif, ce sont les migrations urbaines qui sont à décrire –, d'autres dimensions (professionnelles, familiales, etc.) doivent être ajoutées, notamment du fait du rôle important que leurs caractéristiques peuvent jouer pour expliquer les faits résidentiels. Dans la Figure 19, on crée donc une deuxième trajectoire thématique (ici, *professionnelle*) qui est automatiquement liée à la partie de l'ontologie précédemment créée. Au final, l'ontologie produite par le module est un modèle de données apte à stocker des données sur des trajectoires de vie : *i*) observables selon différents points de vue (*i.e.* les différentes trajectoires thématiques définies) et *ii*) incluant une dimension explicative à travers la notion de *facteurs explicatifs*.

À l'issue du processus de modélisation décrit précédemment, le module *Ontology Manager* transmet le modèle au module *Triplestore Manager* qui réalise à proprement parler l'implémentation de l'ontologie dans une structure de stockage dédiée⁸⁷ (voir Figure 18). Le module *Ontology Manager* prend également en charge la configuration du code des API des autres modules pour les rendre compatibles avec l'ontologie de trajectoire de vie créée. Ainsi, par exemple, le module *Collect Manager* est proposé dans une version complètement adaptée pour permettre l'acquisition des données alimentant le modèle.

3.4.2 Acquisition de données

Une fois l'ontologie de trajectoire de vie créée conformément aux besoins du cas d'étude, nous proposons de faciliter la collecte des données des trajectoires de vie à travers la mise à disposition d'un ensemble de programmes d'enregistrement de données automatiquement paramétrés pour peupler le modèle élaboré à l'étape précédente. Ces programmes peuvent être invoqués pour traiter des sources de données fournies sous la forme de fichiers structurés selon nos spécifications et alimentés par le

⁸⁷ La structure de stockage est un *triplestore* GraphDB (<http://graphdb.ontotext.com>), une base de données conçue pour stocker et récupérer un grand nombre de données au format RDF.

traitement de données résultant d'entretiens menés auprès d'un panel d'individus, ou qui proviendraient d'un moissonnage de réseaux sociaux.

Episode résidentiel

Date d'emménagement: 05/2013

Nombre de pièces: 3

Espace extérieur: Jardin, Terrasse

Type de logement: Appartement

Situation résidentielle: Locataire

Loyer: 600 - 900 €

Rechercher un lieu: Ile Verte, Secteur 2, Grenoble, Isère, Au...

Adresse: _____

Code postal: 38000

Ville: Grenoble

Pays: France

Quelles raisons ont influencé votre choix ?

1 Evénement professionnel

+ Ajouter un événement

Précédent Suivant Annuler Valider

Figure 20 - Exemple d'interface développée pour la collecte de données des trajectoires résidentielles
Auteur : Alan Chakma, stage de Master 2.

L'API du module *Collect Manager* peut aussi être exploitée par une application dédiée à destination d'utilisateurs finaux. Dans ce cas, des développements *ad hoc* sont à réaliser pour proposer des interfaces utilisateurs adaptées. La Figure 20 illustre une proposition de formulaire développé dans le cadre de nos premiers tests. Il s'agit ici de récolter des informations relatives à la trajectoire

résidentielle d'individus. La partie supérieure de la figure montre les différents attributs qu'il convient de compléter pour décrire un épisode résidentiel. Les champs tels que *Date d'emménagement* ou encore *Nombre de pièces* sont, par exemple, des attributs qui ont servi à paramétrer le processus de génération de l'ontologie de trajectoire résidentielle décrit dans la section précédente. La figure illustre également la possibilité offerte de choisir la granularité (voir section 3.2.2) à laquelle est donnée la date d'emménagement (ici le mois). Pour l'attribut *Situation Résidentielle*, un vocabulaire contrôlé est par exemple employé (celui de l'INSEE).

La localisation géographique du lieu de résidence est réalisée au moyen d'une carte dotée de fonctionnalités de recherche textuelle pour faciliter la saisie : par exemple, lorsque le nom d'un quartier est donné (comme ici « Ile Verte »), la carte se centre sur le polygone correspondant et permet de zoomer si besoin pour donner des informations plus précises (place, rue, etc.). Enfin, la partie basse de la Figure 20 montre comment l'utilisateur peut saisir des facteurs explicatifs liés au déménagement (événement) qui marque le début de l'épisode décrit : il peut choisir un type d'événement dans la liste (ici seul est visible un événement de la thématique familiale, mais il pourrait s'agir d'un événement professionnel) et renseigner les champs associés qui deviendront alors accessibles.

3.4.3 Enrichissement des données collectées

Lors de cette phase, des ressources externes sont exploitées pour enrichir les données afin de faciliter et d'augmenter les possibilités d'analyse ultérieures des trajectoires. Il s'agit essentiellement à cette étape de compléter la représentation des connaissances relatives aux différents éléments constitutifs de la trajectoire, par la création de prédicats supplémentaires dont les sujets sont les concepts de l'ontologie de trajectoire de vie et les objets sont des concepts issus d'autres ressources du Web des LOD.

Nous avons évoqué en section 3.3.1.2 que la dimension géographique d'une trajectoire est établie par le recours à des objets de type *Geo:Feature* défini dans GeoSPARQL, ce qui permet d'exploiter notre approche pour la gestion de territoires évolutifs. Plus exactement, c'est ici la capacité à faire référence aux différentes versions d'une même unité territoriale qui est utilisée : peu importe la façon dont est désignée (nommée) cette unité au moment de la collecte puisque les graphes élaborés selon les ontologies TSN et TSN-Change permettent de s'affranchir des évolutions de désignation (nommage) au cours du temps. En d'autres termes, même si une localisation n'est pas nommée à l'identique par deux individus l'ayant cependant tous deux fréquentées (à des temps différents), il est possible de déduire ce fait. Par exemple, en 1992, la commune de Blaiserives en Haute-Marne devient Doulevant-le-Château⁸⁸ ; notre approche permet de retrouver les individus ayant séjournés sur ce seul et même territoire avant et/ou après le changement de noms (dont on ne pourrait pas deviner qu'ils concernent la même commune). Cela permettra aussi de ne pas conclure de manière erronée au déménagement d'un individu sur la base d'un changement de nom d'unité territoriale.

L'ouverture sur le LOD permet de contextualiser (sociologiquement, politiquement, économiquement, etc.) les informations de trajectoires. Ainsi, en exploitant l'immense base de données distribuée que constituent les LOD dans le Web, les descriptions de certains éléments de notre ontologie de trajectoire peuvent être enrichies par la recherche d'autres ressources apportant des compléments d'information utiles pour les analyses, à des échelles spatiales variées ou selon divers angles d'étude. Ainsi, si les lieux de résidence, comme des villes, explicitement mentionnés lors de la collecte, peuvent être mis en relation avec un tout autre maillage thématique⁸⁹ comme celui concernant les bassins d'emploi (ou encore les cartes scolaires), à des fins d'études des migrations résidentielles intégrant une dimension économique (ou sociétale). Dans le même ordre d'idée (et comme introduit en section 3.3.2.3), les facteurs explicatifs peuvent aussi référencer de événements de divers ordres dont la description est disponible dans le Web des LOD.

⁸⁸ Décret du 19 octobre 1992, publié au *Journal officiel* n° 247 du 23 octobre 1992, page 14771 <https://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=INTA9200451D>

⁸⁹ sous réserve bien sûr qu'une représentation de ces données existent dans le LOD.

3.4.4 Exploitation des données de trajectoires

Des fonctionnalités d'exploration et d'analyse des données de trajectoires de vie sont préprogrammées pour interroger les données. Accessibles via le module Exploitation Manager, ces fonctionnalités prennent la forme de requêtes paramétrées permettant d'obtenir des résultats conformes à des critères spatiaux, temporels et/ou thématiques exploitant l'ontologie de trajectoire de vie. Des interfaces de restitution et d'exploration visuelle adaptées (incluant des cartes, graphiques, cube spatio-temporels, *etc.*, éventuellement combinés et synchronisés) peuvent être développées en ayant recours à ces fonctionnalités (via l'API). Nous donnons ci-après quelques exemples de ce qu'il est possible d'obtenir (voir (Noël *et al.*, 2016) pour des exemples de requêtes exprimées en SPARQL).

Concernant la dimension temporelle, les requêtes permettent, par exemple, de connaître la situation d'un (ou plusieurs) individu(s) à un instant t ou pendant un intervalle de temps $[t_1, t_2]$. Par situation, nous entendons la restitution des épisodes et événements observés au temps t ou sur $[t_1, t_2]$. Ici, nous ré-exploisons et adaptions l'approche identitaire utilisée dans les travaux de C. Bernard (voir pour mémoire, Équation 1 – Calcul de similarité entre deux unités appartenant à des versions différentes de TSN, page 29). Par analogie, l'approche peut être utilisée pour déterminer si deux épisodes d'une trajectoire de vie sont assimilables à un seul et même épisode, en tolérant une certaine variabilité des valeurs d'attributs descriptifs de l'épisode. Au moins deux cas d'application peuvent être trouvés à cela :

- réduire le nombre d'épisodes d'une trajectoire de vie en ignorant des modifications jugées mineures en regard des analyses à réaliser
- comparer des épisodes de trajectoires appartenant à des individus différents avec un seuil de tolérance sur tel ou tel critère pour décider ou non de leur similarité.

En considérant une entrée spatiale, il est possible de connaître les lieux qui font partie de la trajectoire des individus mais surtout de les mettre en regard, pour analyser par exemple la localisation des résidences (trajectoire résidentielle) et des lieux de travail (trajectoire professionnelle) et évaluer l'impact des proximités géographiques dans les choix des individus. Les localisations sont interrogeables et restituées selon la granularité retenue pour l'acquisition des valeurs des attributs spatiaux du modèle, ou à des niveaux supérieurs pour des analyses plus macroscopiques, en exploitant la hiérarchie des concepts. Cette notion de navigation dans la hiérarchie des concepts est tout aussi valable pour des critères temporels et thématiques. Il s'agit d'exploiter les relations d'inclusions hiérarchiques de concepts pour donner à voir l'information à différents niveaux d'agrégations. La réalisation est inspirée des travaux de (Silva *et al.*, 2017) dont l'approche, reposant sur la théorie des granules (Bittner and Smith, 2003), offre des représentations de phénomènes à différents niveaux de détails.

Le modèle permet également de connaître les facteurs explicatifs rapportés par les individus pour éclairer leurs trajectoires. Il est par exemple possible de s'intéresser à un événement précis de la trajectoire et d'en rechercher les causes en parcourant le modèle. Plus généralement, on peut rechercher tous les événements d'une thématique donnée ayant été totalement influencés par des événements d'une autre thématique. Des relations de causalité plus complexes peuvent être identifiées, par exemple découvrant des récurrences dans les enchaînements d'événements.

Naturellement, des requêtes alliant dimensions temporelles, spatiales, thématiques, incluant des critères relatifs à des facteurs explicatifs sont possibles. Ramenées à l'objectif d'observation des migrations urbaines et d'aide à la décision dans les domaines de l'urbanisme et de l'aménagement du territoire, de telles possibilités de requête permettront de vérifier des hypothèses qu'il convient de formuler en collaboration avec des spécialistes de ces enjeux.

Nous avons indiqué en section 3.1.1 que si répondre à la question « pourquoi les individus déménagent-ils ? » contribue à mieux comprendre les choix résidentiels, il ne faut pas pour autant négliger la question duale : « pourquoi ne déménagent-ils pas ? » (Coulter and van Ham, 2013). Pour traiter ces questions, l'importance de prendre en considération les aspirations des individus est soulignée par (Coulter and van Ham, 2013) afin de pouvoir mettre en relation une trajectoire effective (ou avérée) et une projection vers le futur de l'individu. Notre approche peut permettre cela dès lors qu'on considère ces « aspirations » comme des trajectoires virtuelles représentables par application de nos patrons. Au lieu de retranscrire des observations avérées, le modèle est alors exploité pour représenter les projets des individus quant à leurs trajectoires de vie. Les concepts présentés

précédemment (épisodes, événements et facteurs explicatifs) sont applicables dans ce but comme nous l'avons montré dans (Noël *et al.*, 2016). Nous pouvons ainsi représenter la projection que font / que faisaient les individus de leur situation pour le futur. La notion de *futur* est en effet ici relative : retracer la trajectoire de vie d'un individu afin de l'analyser nécessite de s'intéresser à son passé et aux aspirations qu'il avait à cette époque-là (*i.e.* ce qui constituait au temps *t* son futur).

3.5 Synthèse et prolongements

La collaboration à caractère pluridisciplinaire à l'origine de ces travaux vise, dans une dimension opérationnelle, à concevoir et développer un outil informatique support d'une approche d'observation des migrations urbaines et/ou péri-urbaines, d'identification des logiques de résidence et de déplacement, la mise en exergue des dynamiques métropolitaines, le tout à des fins d'aide à la décision dans les domaines de l'urbanisme et de l'aménagement du territoire. Il n'existe pas à ce jour d'outil à destination des habitants permettant de collecter, sur le mode participatif des données relatives à leurs trajectoires de vie. Il n'existe de fait pas non plus d'outils d'exploration avancée permettant, à des experts sociologues, urbanistes ou du domaine de l'aménagement du territoire, de visualiser et d'analyser, selon différentes dimensions (personnelles, professionnelles, socio-culturelles, *etc.*) ces données sur les trajectoires résidentielles. Conjuguer aisément contribution des habitants et analyse des experts dans le domaine des migrations urbaines est un des enjeux visés par l'approche que nous avons présentée. Les premiers développements allant en ce sens ont été évoqués dans ce document.

Par le biais d'une enquête visant à collecter des données auprès d'une trentaine d'individus, nous avons testé la bonne compréhension des propositions faites au niveau conceptuel pour « raconter » une trajectoire de vie à base d'épisodes, d'événements et de facteurs explicatifs. Le bilan évoque une appréhension aisée des concepts mais soulève la question de l'investissement demandé⁹⁰ pour retracer toute une vie précisément, et d'autant plus qu'il y a d'épisodes et de facteurs explicatifs à évoquer et que ceux-ci nécessitent de remonter loin dans le passé.

Du point de vue de nos travaux de recherche, nous avons opté ici pour une description moins technique que ce fût le cas en comparaison du chapitre 2, mais l'ensemble des principes et de bonnes pratiques adoptés pour TSN et TSN-Change restent ici valables pour la construction d'ontologies dont la vocation est d'être publiées dans le Web des LOD. L'accent a été mis sur la proposition d'une démarche générique à base de patrons, qui, par applications successives de ceux-ci, permet de répondre à un objectif de représentation de connaissances relatives à des trajectoires sémantiques multidimensionnelles (*i.e.* pouvant être décrites selon différents points de vue) et annotées par des facteurs explicatifs de différentes natures (internes ou externes à une trajectoire personnelle, liés à d'autres événements ou à des caractéristiques d'épisodes de vie). Nous avons montré comment cette approche est supportée d'un point de vue méthodologique et logiciel afin de garantir sa mise en œuvre de manière relativement aisée. Ceci répond au cadre applicatif à l'origine de ce travail, qui implique des experts de disciplines telles que la sociologie ou l'urbanisme, pour lesquels un tel outillage vise à faciliter la collecte et l'exploitation de données pertinentes. Un des enjeux qu'on ne peut ignorer à ce stade relève de la capacité à collecter des données, de façon suffisamment massive pour que les analyses menées (que ce soit à des fins de recherche ou de prise de décision) soient probantes.

Il convient donc de s'attarder sur les moyens qui permettront de garantir l'alimentation du modèle. Ceci peut d'autant plus se révéler être une difficulté qu'on peut rapidement tendre vers une certaine complexité du modèle (en multipliant les thématiques, les attributs descriptifs de celles-ci, *etc.*). A partir des développements réalisés pour acquérir des premiers jeux de données (voir section 3.4.2), nous avons aussi expérimenté que la logique avec laquelle les individus souhaiteraient pouvoir « raconter » leurs trajectoires diffère : le récit peut être orienté par la thématique (décliner toute sa vie personnelle, puis professionnelle, *etc.*), purement chronologique (voir ante-chronologique) ou encore suivre un cheminement que l'on pourrait qualifier de guidé par les facteurs explicatifs. On comprend ainsi aisément qu'une application d'acquisition des données peut rapidement ne plus être un simple cadre mais devenir un véritable carcan qui se révèle être un frein à la participation des contributeurs.

Deux pistes peuvent ici être suivies qui concernent la conception et le développement d'outils adaptés, simples et utilisables dédiés aux contributeurs. Au delà d'une réflexion sur les bonnes

⁹⁰ investissement à rapprocher de l'envie de le faire mais aussi de l'effort cognitif que suppose se remémorer les faits.

propriétés des interfaces proposées (intuitive, efficace, *etc.*), une première voie est celle de la mise en exergue des paramètres qui susciteront la participation des contributeurs. Autrement dit, que faut-il apporter ou offrir à l'individu pour l'amener à participer ? Si, pour certains utilisateurs-contributeurs, la cause scientifique peut suffire à motiver la participation, d'autres ne seront enclins à participer que si un intérêt plus personnel est satisfait. Un travail exploratoire sera donc mené sur ce point, dans le cadre de notre collaboration avec des sociologues, pour dresser une typologie des attentes. Une des pistes à envisager est celle consistant à offrir aux contributeurs, en contrepartie de leur participation, des outils de scénarisation de leur propre trajectoire de vie. C'est, à titre d'exemple, le principe adopté dans le projet *Memory Traces*⁹¹ réalisé par le MIT Design Lab en partenariat avec le consulat italien de la ville de Boston, à l'occasion du 150^{ème} anniversaire de l'unification de l'Italie. Des chercheurs du MIT Design Lab ont recueilli les souvenirs de personnalités de la communauté italo-américaine de la ville de Boston et ont mis à disposition 150 épisodes de vie qu'il est possible d'explorer *via* une application web et une application mobile *in situ*. Cette approche exploite l'intérêt trouvé par certains d'exposer publiquement leurs données personnelles, mais des propositions de diffusions des données scénarisées à des cercles privés pourraient être envisagées⁹².

Une seconde voie à suivre est également en lien avec ces questions de publication de données personnelles sur le web. Il s'agit d'exploiter l'interopérabilité avec les réseaux sociaux tels que *Facebook*, *LinkedIn*, *etc.*, car ceux-ci couvrent de fait des champs informationnels pertinents en matière de trajectoire de vie. L'hypothèse formulée ici est que, si l'on est capable d'extraire et de formater automatiquement des éléments de trajectoires contenus dans les réseaux sociaux pour les rendre compatibles avec notre modèle, alors on réduit d'autant la charge incombant à un contributeur. Il est cependant ici encore évident que les questions de confidentialité des données et de la protection de la vie privée ne pourront être ignorées.

⁹¹ Le site web dédié ne semble plus accessible mais une présentation du projet est visible à l'adresse <https://design.mit.edu/projects/memory-traces>

⁹² Aborder ces questions pourrait être l'occasion de réinvestir, dans le contexte du web des LOD, la question des modèles de contrôle d'accès que J. Gensel et moi avons abordé en 2002 dans le cadre du DEA d'Informatique de Mehdi Adda, intitulé *Modèles de Sécurité pour les Systèmes d'Information basés sur le Web*, DEA Informatique, Système et Communication de l'Université Joseph Fourier, Grenoble.

« Ce que nous appelons le hasard n'est et ne peut être
que la cause ignorée d'un effet connu. »

Voltaire (1694 - 1778)

Récits de situations de crise

4

4.1 Contexte des travaux

4.1.1 Motivations et Enjeux

Menée dans le cadre d'une convention CIFRE avec la SNCF (pôle *Ingénierie et Projet*, département *Lignes, Voies et Environnement (I&P – LVE)*), la thèse de C. Saint-Marc [TH-4] se situe dans le domaine de la géovisualisation et porte sur l'élaboration de méthodes de restitutions adaptées à l'analyse des impacts des inondations sur le système ferroviaire.

Ce travail a conduit à quatre contributions. La première est la formalisation des récits d'inondations dans une ontologie, appelée IDISFER (*Impacts Des Inondations sur le Système FERroviaire*), qui décrit à la fois les typologies d'événements issus d'inondations, les effets dominos potentiels et leurs impacts sur le système ferroviaire, et les mesures de réaction pour ramener le système à l'état d'équilibre. En cela, il s'agit d'une proposition à la frontière des ontologies dites de domaine et métier. IDISFER est aussi une ontologie d'application qui vise à être instanciée pour représenter des situations de crise déclenchées par une inondation. À partir de l'instanciation de cinq inondations historiques variées, la deuxième contribution consiste en des principes sémiologiques généralisables pour restituer les récits d'inondation dans leurs dimensions spatiales et temporelles et en montrant les effets des relations de causalité. La troisième contribution est l'intégration de ces propositions dans une interface de géovisualisation, incluant des visualisations innovantes et alternatives pour mieux appréhender les temporalités associées aux événements. Cette interface de géovisualisation a fait l'objet d'une expérimentation auprès des experts du domaine ferroviaire. Grâce à l'analyse des résultats et du déroulé de l'expérimentation, la quatrième contribution consiste en un modèle de protocole expérimental réutilisable, adapté au test d'interfaces de géovisualisation.

Dans ce manuscrit, nous ne présentons qu'une partie des travaux menés dans le cadre de cette thèse en privilégiant une vision qui s'inscrit dans la continuité des recherches présentées dans les deux chapitres précédents puisqu'il est, là encore, question de représenter une évolution. Nous explorons, dans ce chapitre, *l'évolution d'une situation* qui débute par un événement, voit les différentes répercussions de cet événement impacter le territoire et ses objets, nécessite que des mesures soient prises en réaction et ce, jusqu'à un retour à la normale. Ce travail est également l'occasion de montrer comment s'articulent nos deux principales thématiques de recherche : la représentation des connaissances et la géovisualisation. Plus qu'une juxtaposition de deux « mondes », nous les envisageons en synergie. À l'occasion de la rédaction de ce document, nous montrons comment l'approche ontologique peut être exploitée pour formaliser la connaissance portée par les principes sémiologiques généralisables évoqués plus haut, en vue de leur exploitation dans un environnement de géovisualisation. Nous prolongeons ainsi les travaux menés avec C. Saint-Marc, en posant les premiers jalons d'une approche allant dans ce sens. Avant de l'aborder en section 4.3, nous revenons sur les motivations au cœur de la thèse de C. Saint-Marc.

Le *Retour d'Expérience* (REX) est une approche de maîtrise du risque appliquée suite à un événement dommageable : « *il permet de comprendre un enchaînement de faits ayant conduit à un événement redouté et d'en dégager des recommandations d'amélioration ou des bonnes pratiques en termes de conception et de pilotage du système ou de la crise* » (Chapurlat and Aloui, 2007). Les étapes principales de la réalisation d'un REX sont : 1) la collecte de données sur l'événement dommageable, et 2) la rédaction d'une note de synthèse listant les recommandations et les bonnes pratiques à mettre en œuvre. C. Saint-Marc (Saint-Marc, 2017) souligne l'importance de l'approche pour la SNCF, dans son rôle de gestionnaire d'infrastructure ferroviaire et dans le cas particulier des inondations⁹³:

- garantir la sécurité et la qualité du réseau ferré pour les opérateurs ferroviaires. Les REX réalisés à la suite d'inondations participent ainsi à l'établissement de règles de dimensionnement des ouvrages hydrauliques (Chazelle et al., 2014), ou encore, permettent d'installer des dispositifs de détection aux points du réseau identifiés comme étant vulnérables et d'instaurer des protocoles de réaction lors de la survenue d'un événement menaçant (par exemple, l'arrêt de l'ensemble des circulations pour vérifier la sécurité des voies).
- limiter les coûts liés aux événements d'inondation (coûts des travaux de réparation, de maintenance, ou encore coûts liés à l'interruption des circulations commerciales). Ici, les REX sont des sources d'informations précieuses sur les points du réseau sensibles aux inondations, et sur les procédures de gestion de crise mises en place dans le passé : lieux de garage du matériel roulant, itinéraires de substitution, solutions de transports alternatives pour les voyageurs, etc.

Les enjeux d'une exploitation efficace de toute donnée pouvant contribuer à ces objectifs sont donc importants et les REX, bien qu'utiles, présentent certaines limites. Un REX prend, en effet, souvent la forme d'un dossier intégrant plusieurs données, généralement aux formats tableur et textuel, organisées par métier ferroviaire (par exemple, REX ouvrage d'art, REX ouvrage en terre, REX voies et abords, REX gestion de crise, etc.). Or, comprendre la situation de crise dans sa globalité, tout en cernant l'ensemble des causes et des impacts à chaque étape, ferait sens. Par ailleurs, une perception fine de la situation à travers les dimensions spatiale et temporelle des événements n'est pas favorisée par les modes de restitution traditionnels. Des solutions valorisant la connaissance relative à l'ancrage des événements dans le territoire et à leur chronologie, sont utiles pour comprendre leur enchaînement et en tirer des enseignements.

Dès lors, notre approche vise, au final, un outil de géovisualisation dédié aux récits d'une situation de crise, depuis son événement déclencheur jusqu'au retour à la normale, en passant par les actions de gestion de la situation impliquant différents acteurs. Un tel outil augmenterait les REX habituels par des contributions relevant à la fois du fond et de la forme. Sur le fond, l'accent est mis

⁹³ Le risque naturel inondation est au cœur de la problématique de la thèse de C. Saint-Marc. Il s'agit d'un des risques à fort enjeu pour la SNCF qui dispose par ailleurs de très nombreux documents d'archives sur le sujet, exploités dans les recherches de C. Saint-Marc. À cette occasion, une contribution méthodologique pour la collecte a été réalisée en collaboration avec D. Cœur, historien. Elle est présentée dans (Saint-Marc, 2017), *Annexe 1 : Méthode de collecte des documents d'archives*.

sur la nécessité de couvrir l'ensemble de la situation, à la fois dans son emprise spatiale et dans sa durée, mais aussi dans sa dimension inter-métiers, afin de mettre en exergue l'impact systémique du phénomène marquant le début de la situation. Dans la forme, des techniques de géovisualisation innovantes sont proposées pour rendre compte visuellement de cette dimension systémique en insistant sur la spatialité, la temporalité et les relations de causalité des événements ayant touché tous les composants de ce système.

Comme déjà évoqué, donner à voir une information, dans une forme qui favorise la compréhension de ce que porte en elle cette information, a comme prérequis de disposer de cette information sous une forme structurée et associée à une sémantique riche. Le premier axe du travail nécessitait de s'intéresser à la représentation des connaissances relatives aux situations de crise résultant d'inondations et touchant le système ferroviaire. Les approches existantes explorées à cet effet sont mentionnées dans la section suivante (voir section 4.1.2), avant que soit exposé le positionnement de notre proposition (voir section 4.1.3).

4.1.2 Représentations des risques naturels et de leurs effets

Les ontologies dédiées à la représentation des risques naturels décrivent souvent cinq aspects complémentaires (Bénaben et al., 2008; Hajji, 2005; Kruchten et al., 2007; Scheuer et al., 2013) :

- Le territoire physique soumis au risque (à travers la topographie, l'hydrologie, *etc.*),
- Les éléments exposés (*e.g.* populations, infrastructures, activités, *etc.*)
- Les phénomènes naturels et leurs propriétés telles que l'intensité, la probabilité, l'étendue, *etc.*
- L'impact des événements naturels sur les éléments exposés (*i.e.* les dommages), et
- Les acteurs, procédures, ressources et territoires de la gestion de crise.

Etant donné le contexte institutionnel de la thèse de C. Saint-Marc une attention particulière a été portée à la description des infrastructures et des ressources qu'elles portent ou transportent, ces concepts étant au cœur du système ferroviaire et particulièrement vulnérables aux inondations.

Les ontologies génériques décrivant une infrastructure considèrent le plus souvent qu'elle est composée d'éléments, qui peuvent être connectés entre eux comme, par exemple, dans (Kruchten et al., 2007). Dans cette proposition, chaque élément est défini par un *état*, qui reflète sa capacité à remplir son rôle, et par un *statut opérationnel* (*i.e.* en service ou hors-service), qui dépend de son état et de l'état des éléments qui lui sont reliés. Les relations entre éléments d'infrastructure sont soit une relation de dépendance (si A est hors-service, alors B est hors-service), soit une dépendance avec délai (si A est hors-service, alors B sera fatalement hors-service quelques temps après) ou une co-location (A et B partagent le même emplacement physique – par exemple, une fibre optique et un tuyau de gaz qui passent dans un pont routier). Scheuer et al. (Scheuer et al., 2013) modélisent l'interdépendance entre diverses infrastructures qui sont physiquement connectées grâce à une matrice d'incidence qui décrit les effets de chaque élément *i* sur tous les éléments *j*, au cours du temps. Les auteurs de (Conrad et al., 2006) décrivent dans leur modèle les effets en chaîne potentiels si l'un des quatre systèmes qu'il considère (quatre types de réseaux d'infrastructure : l'énergie, l'eau, les télécommunications et les transports) est en défaut.

Le cas particulier du système ferroviaire fait l'objet de différentes ontologies. Certaines adoptent un niveau de description très détaillé visant à répondre à des questions spécifiques, telles que les ontologies sur l'interaction véhicule-voie (Lu et al., 2006), sur le trafic ferroviaire britannique (*cf.* *Linked Railway Data Project*⁹⁴), les horaires des trains (Raja Mohan and Arumugam, 2011) ou encore la validation de nouveaux plans de lignes (Lodemann and Luttenberger, 2010). D'autres ontologies optent pour un niveau systémique global. Nous nous concentrons sur ces dernières, d'une part parce que les inondations sont susceptibles d'affecter tous les composants du système ferroviaire, et d'autre part, parce que les informations décrivant les impacts des inondations historiques sur les composants ferroviaires sont généralement disponibles à un niveau de détail relativement faible (par exemple, « rupture du rail », plutôt que « rupture de la vis maintenant le rail ») (Saint-Marc, 2017). La Railway Domain Ontology (RDO) développée dans le projet InteGRail (Umiliacchi et al., 2008) est la première ontologie à décrire le système ferroviaire dans son ensemble, à un niveau macro. Elle reprend et

⁹⁴ <http://ontologi.es/rail/#project>

formalise les concepts du standard d'échange de données RailML⁹⁵. Le projet RaCoOn (Railway Core Ontology) (Morris et al., 2015) vise à modéliser le domaine ferroviaire avec plus de détails que la RDO, en combinant un jeu d'ontologies du domaine ferroviaire.

Dans notre contexte, les composants du système ferroviaire sont susceptibles d'être impactés par les risques naturels. La description des impacts des phénomènes naturels sur les infrastructures est donc explorée, en considérant les ontologies qui décrivent les événements dans le cadre des risques⁹⁶, et les effets dominos qui les lient.

Kruchten et al. (Kruchten et al., 2007) distinguent deux types d'événements : les *événements catastrophiques*, qui sont extérieurs aux objets dont ils affectent l'état, et les *événements internes*, qui reflètent des changements « spontanés » d'état des objets (les dégradations et réparations sont classées dans cette seconde catégorie). Dans leur modèle orienté réseaux, un événement peut affecter l'état des infrastructures, mais aussi le territoire qu'elles desservent (population, activités, région). Il est donc possible de catégoriser les événements issus de risques selon leurs causes ou leurs effets. Dans l'ontologie proposée dans (Provitolo et al., 2009), les événements à risque engendrent des dommages et les événements, tout comme les dommages, affectent des éléments du territoire. Les auteurs de (Scheuer et al., 2013) étendent cette conception avec l'objectif de quantifier économiquement les risques. Ils proposent de relier tout élément à risque à chaque phénomène susceptible de l'impacter par une fonction de susceptibilité qui modélise les dommages causés. Elle prend comme entrée un paramètre d'intensité et délivre en sortie un ratio de dommages. Ce résultat sert ensuite à calculer l'impact de l'aléa, qui est le produit du ratio de dommages et de la valeur monétaire de l'élément.

Dans le domaine des risques naturels, l'« effet domino », terme issu du domaine des risques industriels (Heinrich, 1931), désigne le fait qu'un aléa initial entraîne, par réaction en chaîne, d'autres événements. D'une situation initiale de faible impact, un scénario d'ampleur catastrophique peut se dessiner. L'ontologie de (Dubos-Paillard and Provitolo, 2012) décrit les effets dominos et les dommages associés et vise à étudier les catastrophes combinant risques naturels et risques industriels. Dans cette ontologie, la classe *événement* a deux relations avec elle-même : une relation de composition (un sous-événement peut composer un événement plus large) et une relation causale (un événement peut causer un événement).

4.1.3 Positionnement de l'approche

Les ontologies existantes pour la description du système ferroviaire, des risques naturels et de leurs impacts introduisent, séparément, de nombreux concepts : le concept de haut niveau d'événement, la distinction « risque / phénomène / événement », les relations entre composants d'une infrastructure, les dommages ou encore les effets dominos. À notre connaissance, il n'existe toutefois pas d'ontologie de domaine qui décrive, de manière intégrée, les types d'impacts spécifiques des événements issus de risque naturels sur une infrastructure de transports. De plus, plusieurs ontologies décrivent la gestion de crise et l'organisation des acteurs (partie non abordée dans le cadre de la thèse de C. Saint-Marc) mais il ne semble pas exister d'ontologie pour représenter la période post-crise alors que, par exemple, l'étape de travaux sur l'infrastructure, l'adaptation des circulations en situation perturbée ou les événements de la période jusqu'au retour à la normale des circulations sont des informations importantes dans le cadre d'un REX. Enfin, considérer le système ferroviaire à un niveau de détail intermédiaire, distinguant, par exemple, les composants de la voie ferrée : rail, traverse, ballast, est pertinent pour décrire les impacts d'une inondation. Or, les ontologies décrivant l'ensemble du système ferroviaire restent à un niveau macro (par exemple, niveau « voie ferrée », sans distinguer ses composants)⁹⁷.

⁹⁵ RailML décrit 3 volets du système ferroviaire, au format XML : l'infrastructure, le matériel roulant et la gestion des horaires de circulation. <https://www.railml.org/en/>

⁹⁶ Des ontologies génériques conceptualisant les événements existent et peuvent être également utiles pour l'approche. Voir (Troncy et al., 2010) pour une étude comparative et la présentation de l'ontologie intégrative LODE (Linking Open Descriptions of Events). Nous limitons ici la présentation aux ontologies d'événements traitant plus spécifiquement du risque.

⁹⁷ Le projet RaCoOn vise une description plus fine du système ferroviaire mais l'ontologie n'était pas disponible au moment de la réalisation des travaux présentés.

L'ontologie IDISFER (qui décrit les *Impacts Des Inondations sur le Système FERroviaire*) a été élaborée dans le cadre de la thèse de C. Saint-Marc selon la démarche préconisée dans (Uschold and Gruninger, 1996). Elle vise à représenter l'information sur les inondations historiques de manière structurée, en vue, notamment de son exploitation par un outil de géovisualisation. IDISFER se positionne comme un support pour la réalisation de Retour d'Expérience couvrant l'ensemble des domaines métier du système ferroviaire, en modélisant le récit des inondations et de leurs impacts sur ce système représenté à un niveau de détail adapté à l'analyse des événements le touchant.

L'ontologie IDISFER est, d'une part, un outil de restitution de la connaissance métiers relative :

- aux phénomènes naturels qu'on peut observer lors d'une inondation ;
- aux incidents résultant de ces phénomènes et à leurs impacts potentiels sur les objets ferroviaires ;
- aux mesures de réaction que peut engager le système pour prendre en charge ces impacts.

La particularité de cette ontologie est qu'elle permet la mise en exergue des *effets dominos* auxquels on peut ou doit s'attendre entre ces éléments.

D'autre part, l'ontologie IDISFER, en tant que support pour la formalisation des événements observés, offre un cadre formel pour représenter toute inondation ayant touché le système sous la forme d'une succession d'événements (de type phénomènes, incidents et réactions), localisés dans le temps et dans l'espace et liés par des relations de causalités.

Cette contribution est par ailleurs exploitée comme donnée d'entrée dans une approche originale de définition de règles guidant la production cartographique (voir section 4.3).

4.2 Ontologie d'événements issus de risques et leurs effets dominos

4.2.1 Vue d'ensemble sur l'ontologie IDISFER

Cinq cas d'études d'inondations historiques⁹⁸, issues de différents régimes hydro-climatiques, ont été retenus pour constituer un corpus documentaire à partir duquel identifier les concepts à intégrer dans l'ontologie. Au total, 801 documents de diverses natures ont été collectés sur les cinq inondations: rapports techniques des services ferroviaires, feuilles de budget, courriers échangés entre services ou avec des tiers, coupures de presse, photographies, dossiers de travaux, plans et cartes. Une synthèse par inondation a été produite dans un format tabulaire pour retracer le récit d'une inondation : chaque événement survenu pendant l'inondation est répertorié et décrit notamment par son type, par sa localisation, sa date, sa description, sa source documentaire, ainsi que la nature des imperfections éventuellement relatives à chacun de ces attributs⁹⁹. L'analyse de ces données nous a permis d'aboutir à la proposition de l'ontologie IDISFER, dont la Figure 21 donne un aperçu de l'organisation générale.

Les classes principales de l'ontologie distinguent les concepts de *railwaySystemObject* qui correspondent à des éléments de l'environnement ferroviaire et de *genericEvent* qui visent à traduire des événements susceptibles de se produire dans cet environnement. La suite des concepts n'est pas présentée dans des termes ontologies pour alléger le discours.

L'ontologie permet la représentation d'événements qui spécialisent *genericEvent*, de quatre natures : *Phénomène naturel* (*idisfer:NaturalPhenomenon*), *Incident* (*idisfer:Incident*), *Perturbation* (*idisfer:Disturbance*) et *Réaction* (*idisfer:Reaction*). Un *Phénomène naturel* provoque (i.e. induit ou peut engendrer) un *Incident* subdivisé en deux catégories, suivant la typologie utilisée en interne dans l'entreprise en *Incident de 1er ordre* et *Incident de 2nd ordre* (respectivement *idisfer:OriginIncident* et *idisfer:InducedIncident*). Les premiers sont ceux qui peuvent être causés par un facteur extérieur au système ferroviaire (donc un *Phénomène Naturel*), tandis que les seconds découlent uniquement de

⁹⁸ Ces inondations font l'objet de monographies dans la thèse de M. Boudou (Boudou, 2015). Il s'agit de l'inondation du Tarn de mars 1930 (débordement de cours d'eau), l'inondation d'octobre 1940 dans les Pyrénées-Orientales (crue éclair), l'inondation de l'hiver 1947-1948 dans les Vosges (débordement lent amplifié par la fonte nivale), l'inondation de novembre 1999 dans l'Aude (crue éclair), l'inondation de 2001 dans la Somme (inondation lente par remontée de nappes phréatiques).

⁹⁹ La thèse de C. Saint-Marc a également été l'occasion de travailler sur la notion d'imperfection et de contribuer à l'état de l'art par la proposition d'une nouvelle typologie. Ce travail est présenté dans (Saint-Marc, 2017), *Annexe 3: Elaboration d'une typologie des informations imparfaites* et dans (Saint-Marc et al., 2016).

dommages de premier ordre survenus au préalable. Les incidents ont un impact sur les objets du système ferroviaire, eux-mêmes subdivisés en quatre catégories (voir section 4.2.2). Les incidents induisent ou peuvent engendrer des *Perturbations* des circulations – seul type d'*Objets de l'exploitation* concerné par les *Perturbations*, d'où leur mise en exergue dans la Figure 21. Les *Perturbations* induisent la mise en place de mesures de *Réaction*, qui agissent sur le système ferroviaire et peuvent elles-mêmes engendrer d'autres *Perturbations* (e.g. la limitation des vitesses de circulation le temps de réaliser des travaux).

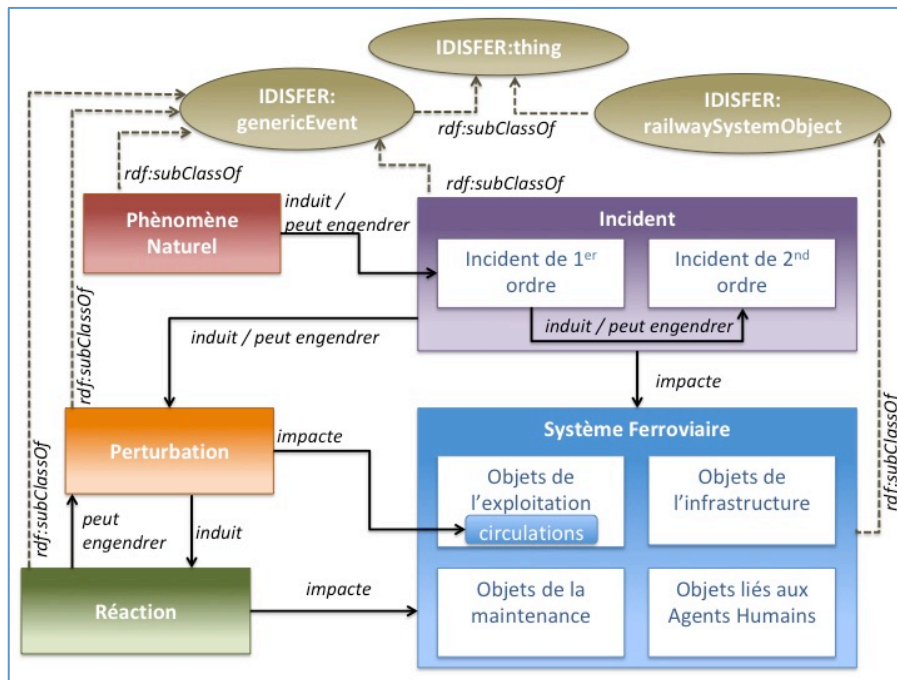


Figure 21 - Modèle simplifié de l'ontologie IDISFER

Adaptation d'un schéma de C. Saint-Marc, 2017.

Les différentes composantes de cette ontologie sont liées entre elles par des relations de causalité et/ou d'impact qui sont à la base des effets dominos. Cette ontologie décrit de telles relations en distinguant deux types : la relation *induit* (*involves*) est employée pour lier deux éléments tels que si le premier se produit, alors le second se produit obligatoirement ; la relation *peutEngendrer* (*mayCause*) correspond à une relation possible mais pas systématique. Présentées dans la Figure 21 à un haut niveau d'abstraction, ces relations sont définies pour les concepts qui participent à la représentation des connaissances relatives à chacune des cinq parties principales de l'ontologie. Les sections suivantes les présentent rapidement.

4.2.2 Aperçu des différentes parties de l'ontologie

4.2.2.1 Ontologie relative au système ferroviaire

Le système ferroviaire est décrit par une ontologie dont, la classe racine est *railwaySystemObject*, et qui compte quatre sous-ensembles dédiés à la représentation des concepts (voir, pour une présentation complète, (Saint-Marc, 2017), *Annexe 2*) :

- de l'infrastructure de la voie, représentée par 68 classes (voie, ouvrages d'art, signalisation,...),
- de l'infrastructure de maintenance, comptant 8 classes (ateliers de maintenance, station d'entretiens,...) :
- de l'exploitation ferroviaire, nécessitant 41 classes (circulations, trains, gares...)
- du système humain (représenté uniquement comme preuve de concept de l'intérêt qu'il y a à gérer cette information, par quelques classes relatives au recensement des informations sur les trajets domicile-travail des agents – voir 4.2.2.5)

Tous ces sous-systèmes sont susceptibles d'être impactés par les inondations sous des formes différentes : endommagement voire destruction pour les infrastructures et les autres bâtiments ou installations liés à l'exploitation, perturbation pour les circulations. Le quatrième sous-système témoigne d'un aspect organisationnel relatif aux agents en charge de la maintenance ou des circulations. Plusieurs situations ont été identifiées dans lesquelles, alors que l'infrastructure elle-même n'a pas été touchée par l'inondation, les circulations de train ont été perturbées à cause de la coupure des axes routiers empruntés par les agents dans leurs trajets domicile-travail.

4.2.2.2 Ontologie relative aux phénomènes d'inondation

La représentation des phénomènes naturels reprend pour l'essentiel les concepts de l'ontologie environnementale de haut-niveau SWEET¹⁰⁰ (vocabulaire de la partie *Phenomena* essentiellement). Des phénomènes naturels spécifiques sont ajoutés : l'embâcle, la fonte des neiges, la marée, ainsi que des phénomènes d'inondation : le débordement (de cours d'eau ou de lac), la remontée de nappe phréatique, le ruissellement intense et la submersion marine, issus de la typologie présentée dans (Lang and Coeur, 2014)¹⁰¹. Au total, 6 classes de premier niveau se spécialisent en une trentaine de sous-classes.

Les phénomènes naturels peuvent être liés entre eux par des relations de causalité : par exemple, la pluie peut engendrer le débordement d'un cours d'eau ou une remontée de nappe phréatique, ou encore, la combinaison d'un vent fort et de la marée haute peuvent engendrer une submersion marine.

4.2.2.3 Ontologie relative aux incidents

Les incidents ferroviaires sont classés en deux catégories, conformément à la nomenclature en vigueur dans l'entreprise :

- les *incidents de 1^{er} ordre* (aussi appelés *incidents origine*) sont causés par des phénomènes hydro-climatiques (e.g. le ruissellement peut engendrer l'érosion d'un remblai) ;
- les *incidents de 2nd ordre* (ou *incidents induits*) sont causés par des incidents origines (e.g. le gonflement d'un talus peut engendrer une déformation linéaire de la voie), et jamais directement par des phénomènes naturels.

Actuellement, 19 classes décrivent des incidents de 1^{er} ordre, 16 classes représentent des incidents de second ordre, et 35 relations décrivent des effets dominos entre incidents (31 relations *peut engendrer* et 4 relations *induit*).

4.2.2.4 Ontologie relative aux perturbations

Les phénomènes naturels peuvent donc avoir pour conséquences des incidents qui peuvent eux-mêmes entraîner des perturbations dans la circulation des trains. Les perturbations peuvent être de trois types (3 classes dans l'ontologie) :

- *défaut de sécurité* qui représente une menace envers la sécurité du personnel, des voyageurs ou des marchandises convoyées ;
- *défaut de régularité* qui traduit un écart par rapport à l'horaire théorique prévu pour la circulation d'un train ;
- *accident* qui désigne tout dommage matériel subi par un train résultant d'un environnement de circulation anormal.

L'apparition d'un *défaut de sécurité* ou d'un *accident* induit la mise en place de mesures de réaction visant à rétablir le système ferroviaire dans son état d'origine.

¹⁰⁰ <https://sweet.jpl.nasa.gov>

¹⁰¹ L'ouvrage « *Les inondations remarquables en France* » a été élaboré dans le cadre de l'application de la Directive européenne relative à l'évaluation et à la gestion des risques d'inondation (Directive 2007/60/CE du 23 octobre 2007). Une partie des informations relatives aux 175 inondations décrites et qui couvrent une période de plus de deux cents ans, de 1770 à 2011, est issue de la Base de Données Historique sur les Inondations (BDHI). Il s'agit d'une application développée au sein de l'équipe STeamer et financée par le Ministère de l'Ecologie, du Développement durable et de l'Energie (Direction Générale de la Prévention des Risques – DGPR). <http://bdhi.fr/>

4.2.2.5 Ontologie relative aux réactions

Les mesures de réaction sont consignées à l'aide de 11 classes, organisées en 3 types principaux, et peuvent consister en :

- une *action sur les circulations* : ralentissement, interruption, annulation ou retenue de trains en gare, mise en place d'itinéraires ou de modes de transports de substitution dans le cas des interruptions de longue durée,
- une *action sur l'infrastructure*, comme des travaux de régénération ou
- une action sur le *système humain* qui peut prendre la forme d'une délocalisation de l'activité des agents lorsqu'ils ne peuvent plus rejoindre leur lieu de travail habituel.

Parmi les mesures de réaction, les dispositifs de gestion de crise (hors du champ d'étude de la thèse de C. Saint-Marc) pourraient être décrits, en s'appuyant par exemple sur des travaux tels que (Bénaben et al., 2008).

IDISFER, en tant que qu'ontologie métier qui met à disposition la connaissance présentée dans cette section, est disponible au format RDF/OWL¹⁰². Elle contient près de 200 classes. Couplée avec des concepts permettant de représenter des événements réels recensés par le passé, elle devient un cadre formel de représentation pour les REX. Nous abordons ce point dans la section suivante.

4.2.3 Exploiter l'ontologie IDISFER pour représenter les situations passées

Les *Phénomènes naturels*, *Incidents*, *Perturbations* et *Réactions* sont décrits en toute généralité dans IDISFER, sous la forme de spécialisation du concept *genericEvent*, qui traduit des événements susceptibles de se produire. Il s'agit donc de représentations à un haut niveau d'abstraction (par rapport à des événements qui se sont réellement produits). Parallèlement, nous avons déjà évoqué l'ontologie LODE (Troncy et al., 2010) qui permet de caractériser un événement par différents attributs (une identification, une datation, une localisation, etc.). En considérant le nouveau concept *inCrisisEvent* comme sous-classe des concepts *genericEvent* et *lodeEvent*, nous créons une structure permettant la représentation des événements observés lors de situations de crises avérées (voir Figure 22). De manière immédiate à travers l'usage de la relation de subsumption, cette représentation :

- d'une part, est conforme aux connaissances métiers portées par IDISFER dans ses 5 grandes parties, et,
- d'autre part, permet, du fait du recours à LODE, l'identification, la localisation spatiale et temporelle (en autres) des occurrences de *Phénomènes naturels*, *Incidents*, *Perturbations* et *Réactions*.

On note toutefois que les relations de type « effet domino » représentées à ce stade sont celles qui ont effectivement eu lieu. La distinction entre les prédicats induit (*involves*) et peut engendrer (*mayCause*) n'est plus pertinente ici. Aussi, pour représenter les relations de causalité avérées entre des instances e_1 et e_2 de *InCrisisEvent*, le prédicat *causes*(e_1 , e_2) est utilisé tel que e_1 est l'événement cause et e_2 l'événement conséquence.

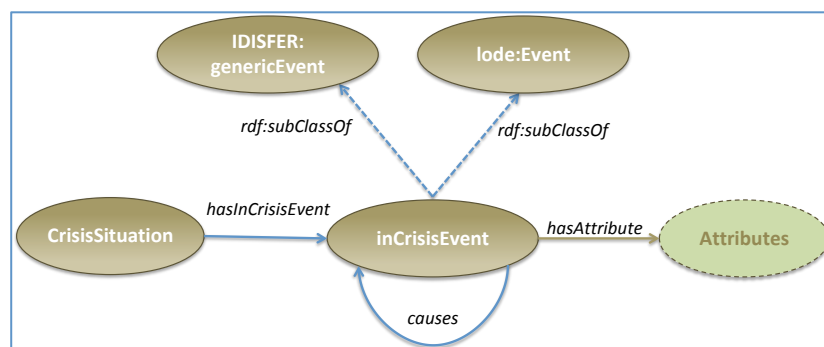


Figure 22 - Illustration simplifiée du concept *InCrisisEvent* dans IDISFER.

¹⁰² <http://lig-membres.imag.fr/saintmar/ontology/IDISFER.html>

De plus, notre approche permet de prendre en compte la notion d'imperfection pour les attributs spatiaux et temporels notamment, grâce à une propriété *aTypeImperfection* pointant vers un terme issu d'une nouvelle typologie que nous avons élaborée à partir de l'étude des propositions produites dans le domaine des sciences de l'information géographique (voir (Saint-Marc, 2017), *Annexe 3*). Ces propriétés assurent la traçabilité des éventuels défauts de qualité présents dans les données (ce qui est assez fréquent dans l'exploitation d'archives). Lors du peuplement de l'ontologie IDISFER, lorsque les dates ou les localisations des événements sont imparfaites ou manquantes, des dates et des localisations possibles peuvent être déduites à partir du contexte des événements.

Le récit d'une inondation, conformément à cette approche, consiste donc à instancier tous les concepts de *genericEvent* selon leur type, en décrivant leurs propriétés par les valeurs adéquates et en établissant les liens de causalité inhérents à leur type et effectivement observés lors de la situation décrite. Deux inondations, l'inondation d'octobre 1940 sur la zone de la vallée du Tech (Pyrénées Orientales) et l'inondation de 2001 dans la Somme, issues de notre jeu d'étude initial et pour lesquelles un travail de préparation des données a été effectué (voir section 4.2.1), ont été instanciées pour vérifier la faisabilité de l'approche et la bonne couverture des situations à l'aide des concepts de l'ontologie. La Figure 23 illustre une partie de la situation décrite pour le cas de 1940.

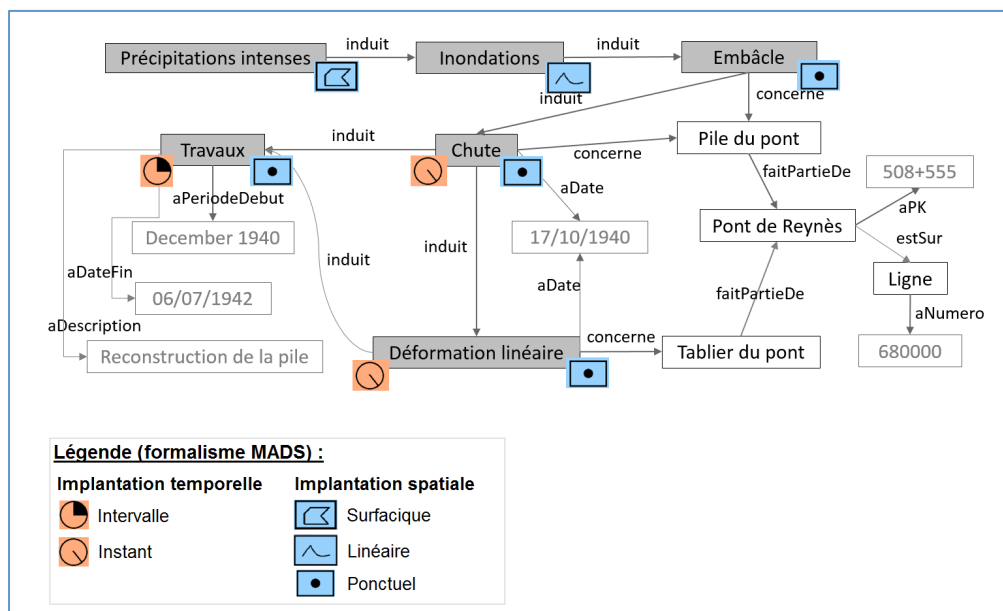


Figure 23 - Illustration d'une instanciation de l'ontologie IDISFER sur l'exemple de l'inondation de 1940 dans les Pyrénées-Orientales

Les termes de l'ontologie (cadres en gris notamment) sont traduits et les indications sur l'implantation spatiale et temporelle des objets utilisent le formalisme MADS (Vangenot et al., 2002; Spaccapietra et al., 2007), ces formes étant plus adaptées pour la diffusion auprès de publics non experts des formalismes ontologiques. Source : Document pour la communication auprès des agents SCNF. Auteur : C. Saint-Marc, 2016.

Le travail réalisé dans le cadre de la thèse de C. Saint-Marc a donc été l'occasion d'aborder une autre démarche de conceptualisation exploitant une approche ontologique pour représenter, cette fois, l'évolution d'une situation. Dans cette présentation, les aspects techniques liés à la représentation des connaissances ont été peu détaillés volontairement. La complexité est moindre, dans l'ontologie IDISFER, pour ce qui concerne les concepts dédiés à la représentation de l'évolution dont nous avons fait notre fil directeur. En effet, l'évolution ne prend ici la forme « que » d'une succession d'événements liés par des relations de causes à effets, s'appliquant en cascade. Pour autant, il ne faut pas négliger la complexité thématique du champ d'application qu'il fallait s'approprier pour produire cette ontologie. Ce travail, mené à partir d'une opération de collecte dans des fonds d'archives auquel a largement contribué C. Saint-Marc, était une condition *sine qua none* pour maîtriser le domaine et faire émerger toutes les relations de type « effets dominos » qui font la particularité de l'ontologie IDISFER. Les contributions de la thèse de C. Saint-Marc portent également sur des propositions en

termes de géovisualisation dédiée aux données représentées dans l'ontologie. Dans la suite de ce chapitre, nous nous appuyons sur ces travaux mais nous prolongeons d'un point de vue méthodologique la réflexion amorcée alors.

4.3 De l'ontologie à la géovisualisation

4.3.1 Contexte, travaux préliminaires et positionnement d'une nouvelle approche

Les nouvelles technologies de l'information et de la communication associées à celles des systèmes d'information géographique, ont permis le développement de nouveaux outils d'exploration et de visualisation des données spatio-temporelles, proposant de nouvelles formes de représentations cartographiques telles que les cartes animées, multimédia ou interactives et introduisant le concept de carte dynamique. En combinant l'animation à l'interactivité, les cartes dynamiques se donnent notamment pour objectif de rendre visibles les temporalités. Tirées par les innovations technologiques, les cartes dynamiques semblent bien accroître les possibilités de visualiser les temporalités, mais leur conception et leur réalisation restent complexes et soulèvent différentes questions de nature conceptuelle ou encore cognitive. Parallèlement, le *Webmapping* est devenu une facette incontournable de la publication de données sous forme de cartes sur le Web qui s'ouvre largement à différents publics, du fait d'une certaine accessibilité des technologies associées (API Google Maps, Leaflet, Openlayers, etc.).

Cependant, la réalisation de « bonnes » interfaces cartographiques n'est pas un processus trivial. Il implique de s'interroger à la fois sur les caractéristiques des données manipulées, sur les relations entre ces données, et sur la façon dont on les organise et les représente visuellement, tout en intégrant les règles du langage cartographique. Ce travail est souvent réalisé de façon plus exploratoire qu'empirique, et avec plus ou moins de succès par des utilisateurs qui souhaitent profiter des atouts en termes d'attractivité des interfaces cartographiques pour diffuser leurs données. Néanmoins, parvenir à un résultat probant relève d'une démarche d'expert et mettre en place des outils et des méthodes capitalisant un savoir-faire expert permettrait d'aider l'utilisateur dans le processus de conception cartographique. Dans le cadre du master de S. Kiv [M2R-1], nous avons fait un premier pas dans cette direction. L'objectif de ce travail était de formaliser la connaissance autour de ce qu'est une interface de visualisation de données géographiques et temporelles, en portant une attention particulière à la partie cartographique. Nous avons notamment travaillé à l'identification de règles qui permettent de déterminer quel type de carte (carte statique, collection de cartes, carte animée, interactive ou dynamique) est approprié compte tenu des données à représenter. Ces règles exploitent notamment le type des données (quantitatives *vs* qualitatives), la temporalité à représenter (durée, fréquence, période de retour, *etc.*), le nombre d'informations à inclure pour identifier la ou les meilleures représentations cartographiques. À partir d'une modélisation UML du domaine d'étude, nous avons proposé une opérationnalisation des connaissances sous la forme d'une ontologie OWL et défini près d'une cinquantaine de règles d'inférence codées en SWRL¹⁰³ (Semantic Web Rule Language). Implémenté sous Protégé, un prototype de ce système d'aide à la décision permettait à un utilisateur de décrire son contexte de production cartographique (type de données, temporalité, *etc.*), à partir de quoi les règles d'inférence permettaient de fournir des recommandations en termes de choix de cartes adaptées au contexte.

La thèse de C. Saint-Marc est pour nous l'occasion de réinvestir ce travail. Présenter les récits d'inondation qui touchent le système ferroviaire, à l'aide de géovisualisations dynamiques, offre de nouveaux potentiels pour les REX. Nous présentons certaines des solutions proposées par C. Saint-Marc (voir (Saint-Marc, 2017), Chapitre VI) allant dans ce sens. Ces propositions s'inspirent de principes éprouvés de la littérature mais apportent également des réponses à certains aux verrous identifiés à l'issue de l'étude de l'état de l'art (Saint-Marc, 2017) (Chapitre II).

Des règles, présentées dans un langage semi-formel, ont été introduites dans (Saint-Marc, 2017) pour expliquer certains principes proposés pour la construction de géovisualisation dédiés aux récits d'inondations représentées dans l'ontologie IDISFER. Dans ce manuscrit, nous donnons une autre vision de ces règles en les immergeant totalement dans une approche ontologique qui part du constat

¹⁰³ <https://www.w3.org/Submission/SWRL/>

suivant : la connaissance contenue dans une ontologie (d'application, métier, ou de domaine) comme IDISFER n'est pas immédiatement transposable dans une interface de géovisualisation comme nous le montrons et l'illustrons à partir du cas IDISFER dans les sections suivantes. Pour donner quelques exemples, il peut être utile d'opter pour une autre organisation des concepts que celle utilisée dans l'ontologie d'origine, pour la simplifier ou donner à voir différemment les concepts, d'une façon plus proche des besoins et de la logique de raisonnement des utilisateurs. Une autre raison peut être liée aux contraintes de la restitution graphique qui ne permettent pas la transposition directe d'un ensemble d'instances (car en trop grand nombre ou dans un format inapproprié).

Dès lors, nous proposons une approche¹⁰⁴, qui permet de prolonger les travaux de C. Saint-Marc dont elle s'inspire, et qui consiste à définir, en des termes ontologiques, les correspondances entre les concepts d'une ontologie et leur représentation exploitée dans une géovisualisation. Nous reprenons en cela certaines des pistes déjà évoquées dans (Fabrikant, 2001). La Figure 24 illustre cette approche.

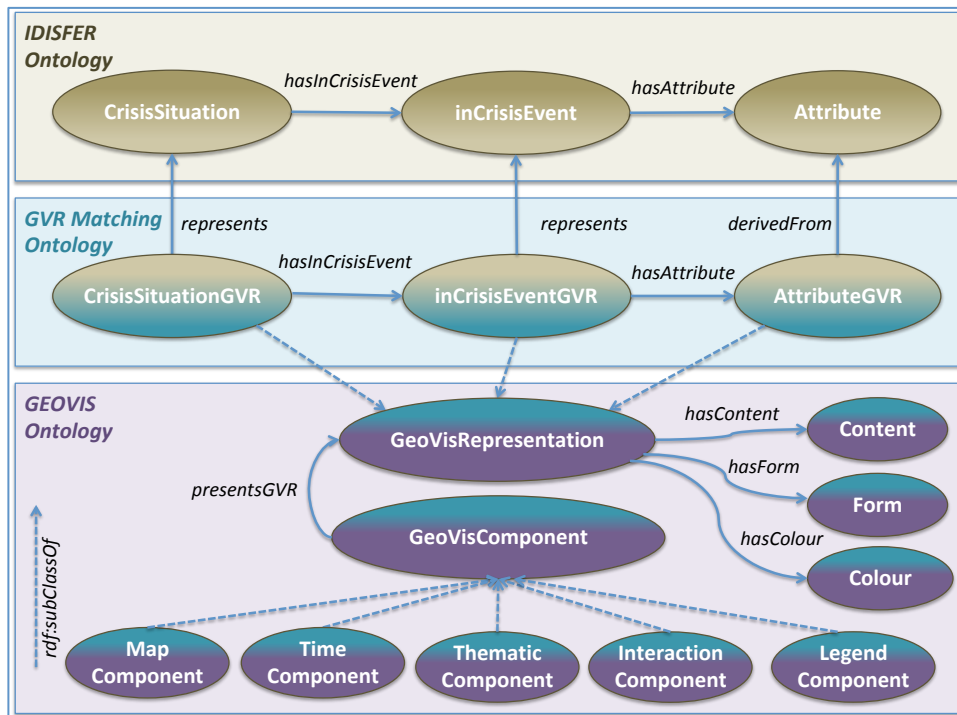


Figure 24 - Schématisation de l'approche GVR Matching

En haut de la figure sont schématisés les concepts de l'ontologie IDISFER qu'on souhaite intégrer dans une géovisualisation (*i.e.* les concepts *inCrisisEvent* et leurs attributs qui portent les informations sur les événements d'inondations d'une situation de crise *CrisisSituation*). L'ontologie *GEOVIS* (schématisée en bas de la figure) donne une représentation des concepts utiles à notre approche : une géovisualisation (concept omis dans le cadre) est constituée de composants (*GeoVisComponent*) qui peuvent être : un composant cartographique (*MapComponent*), un composant temporel (*TimeComponent*), un composant thématique (*ThematicComponent*) et un composant d'interaction (*InteractionComponent*). Un *GeoVisComponent* a vocation à accueillir des représentations géovisuelles (*GeoVisRepresentation*) constituées d'un contenu et d'une forme graphique. Cette ontologie *GEOVIS* n'existe pas à ce jour autrement que dans cette première forme épurée et devra faire l'objet d'un travail de recherche à part entière. Entre l'ontologie *GEOVIS* et l'ontologie *IDISFER* viennent se placer les concepts qui permettent de formaliser la correspondance entre un concept lié à la connaissance (d'application, métier, ou de domaine) et le concept de représentation géovisuelle associé. Ces concepts appartiennent à une ontologie que nous nommons *GVR Matching*

¹⁰⁴ Il s'agit ici de premières réflexions, appliquées à un cas qui s'y prête, et qui seront le support de travaux de recherches à venir. Nous aurions pu présenter ces propositions comme des perspectives de recherche en fin de chapitre, mais il nous paraît plus adapté d'en exposer ici les principes pour mieux les illustrer, et ainsi faciliter la compréhension de l'approche que nous visons.

Ontology (GVR pour *GeoVisual Representation*). Dans la Figure 24, l'événement *idisfer:inCrisisEvent*¹⁰⁵ est représenté par le concept *gvr:inCrisisEventGVR*, sous-classe de *geovis:GeoVisRepresentation*. Le concept *gvr:inCrisisEventGVR* a des attributs désignés par *AttributeGVR* qui sont dérivés (prédicat *derivedfrom*) des attributs associés au concept *idisfer:inCrisisEvent* qui lui correspond. Ceci est défini par :

Soient e une instance du concept *idisfer:inCrisisEvent*
 et $egvr$ une instance du concept *gvr:inCrisisEventGVR*
 telles que $\text{represents}(egvr, e)$ alors
 $\forall a$ tel que $\text{hasAttribute}(e, a)$
 $\exists a'$ tel que $\text{hasAttribute}(egvr, a')$ et $\text{derivedFrom}(a', a)$

La définition de l'ontologie *GVR Matching* vise donc à établir le lien entre les données et leur figuration dans la géovisualisation. Des traitements opérés sur les concepts et instances de l'ontologie IDISFER participent à cette définition : ils se traduisent par l'expression de nouveaux concepts ontologiques portant les *choix de représentation* faits pour visualiser les événements dans l'espace et dans le temps, ainsi que leurs relations de causalité. Dans les sections suivantes nous illustrons cette approche. Les choix de représentations sont basés sur les propositions de (Saint-Marc, 2017).

4.3.2 Prérequis pour une vision systémique des impacts des inondations

Les événements survenus pendant une situation de crise sont nombreux et de natures diverses. Un principe de la cartographie est la sélection des éléments à représenter dans la carte, une densité de données faible favorisant généralement la lisibilité. Néanmoins, la présélection (arbitraire) de données a pour effet d'occulter une partie de la complexité du système soumis au risque, notamment l'interdépendance entre les composants ferroviaires. L'objectif de restitution des effets dominos vise à enrichir l'analyse dans le cadre du REX, et à montrer l'ensemble des événements qui participe de la restitution de la réalité. Aussi, l'occultation d'une partie de la complexité du système est un facteur qui pourrait se révéler contre-productif. Comme l'exprime Le Moigne (Le Moigne, 1999) : « *la simplification du compliqué, appliquée au [système] complexe, a pour conséquence une aggravation de la complexité par mutilation, et non pas la résolution du problème considéré* » (p5).

Ainsi, permettre l'observation des réactions en chaîne entre événements lors d'une situation requiert que tous les événements stockés dans l'ontologie puissent être projetés dans la carte et dans une frise. Cela ne signifie pas pour autant les projeter tous *en même temps* sur la carte ; ainsi des fonctions interactives intégrées dans l'interface de géovisualisation permettent aux utilisateurs la sélection ou le filtrage sur demande, au fil de leurs analyses et en fonction de leurs objectifs. Néanmoins, la projection sur la carte et dans une frise temporelle va nécessiter de régler notamment :

- la question de l'homogénéisation des granularités spatiales et temporelles
- la question de la superposition de la représentation d'événements référencés à une même localisation et/ou une même date
- la question de la représentation de l'étendue spatiale et temporelle, potentiellement large, pour relater la totalité de l'évolution de la situation.

4.3.3 Homogénéisation des granularités spatiales et temporelles

Les événements stockés dans l'ontologie ont des granularités spatiales et temporelles multiples et hétérogènes. Ainsi, des implantations spatiales peuvent être ponctuelles (e.g. une rupture de rail constatée au point kilométrique 508.555), linéaires (e.g. un débordement du cours d'eau du Tech) ou surfaciques (e.g. l'enregistrement d'un record de précipitations sur le département des Pyrénées-Orientales). Les localisations temporelles peuvent être consignées en tant qu'instantanés (e.g. le 17 octobre 1940) ou en tant qu'intervalles (e.g. du 12 janvier au 6 juillet 1940).

¹⁰⁵ *idisfer* est le préfixe utilisé pour désigner l'ontologie IDISFER, *gvr* pour l'ontologie GVR Matching Ontology et *geovis* pour l'ontologie GEOVIS

Co-visualiser dans une même carte des événements ayant ces caractéristiques présente deux verrous principaux : d'une part, il est difficile d'appliquer une symbologie homogène à des objets de géométrie ou de temporalité variables, et, d'autre part, la symbologie doit représenter de façon lisible le fait que ces objets sont de même nature.

Notre proposition est de ramener ces entités à une granularité ponctuelle unique, que ce soit dans le temps ou dans l'espace (Figure 25). Cela est valable dans notre contexte dans la mesure où l'exactitude dans la précision à l'affichage n'est pas requise (ni même toujours possible, du fait de l'imprécision des données collectées...). De plus, il convient de noter que l'information détaillée n'est pas perdue mais reste accessible dans la géovisualisation.

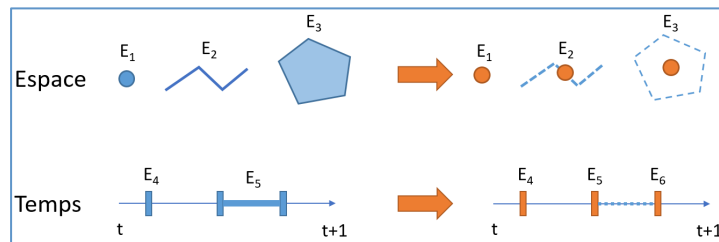


Figure 25 - Méthode de simplification des événements de granularités spatiales et temporelles hétérogènes

Source : image extraite de (Saint-Marc, 2017).

Au niveau spatial, les objets linéaires ou surfaciques sont réduits à leur barycentre (ou le point le plus proche situé sur la polygone ou à l'intérieur du polygone) avant projection dans la carte. La dérivation de l'attribut géométrie est donnée par :

Soient

- e une instance du concept $idisfer:inCrisisEvent$
- et $egvr$ une instance du concept $gvr:inCrisisEventGRV$

telles que $hasGeometry(e, g)$ et $represents(egvr, e)$

avec $g = \{s_1, s_2, \dots, s_n\}$ où les s_i sont les coordonnées des sommets de g

alors on définit $hasGeometry = (egvr, B)$, tel que $B = (x, y) = Barycentre(s_1, s_2, \dots, s_n)$

Au niveau temporel, tous les événements associés à une durée sont scindés en deux événements distincts associés à des instants. Ainsi, un événement « travaux de réparation » de durée définie entre $t1$ et $t2$, est marqué dans la géovisualisation comme un événement *début* (e.g. « *début des travaux de réparation* » associé à la date $t1$) et un événement *fin* (e.g. « *fin des travaux de réparation* » associé à la date $t2$). On pose le principe de dérivation suivant :

Soient

- e une instance du concept $idisfer:inCrisisEvent$ telle que $hasPeriod(e, [t1; t2])$
- et $egvrBegin$ et $egvrEnd$ deux instances du concept $gvr:inCrisisEventGRV$

alors on définit

$represents(egvrBegin, e)$ et $hasInstant(egvrBegin, t1)$ et

$represents(egvrEnd, e)$ et $hasInstant(egvrEnd, t2)$

Les dates sont, de plus, harmonisées à la granularité de la journée, ce qui vaut à la fois pour des horaires plus précis que la granularité du jour (e.g. « à 5h40 le 18 octobre 1940, la voie a été inondée ») et pour des dates imprécises (e.g. « fin des travaux en novembre 1942 »). Pour ces dernières, un jour dans le mois est choisi arbitrairement¹⁰⁶.

¹⁰⁶ Rappelons que le travail de peuplement de l'ontologie IDISFER a été fait à partir d'une analyse des archives disponibles. Si la date n'a pas pu être plus précisément élaborée à cette occasion, alors un choix arbitraire de jour au moment de la restitution graphique est une solution acceptable, et ce d'autant que l'imprécision avérée est bien enregistrée et peut être traduite graphiquement, par le recours à un attribut de forme adaptée.

Soient

- i_1 et i_2 des valeurs de type *owl-time:Instant* ;
 - e une instance du concept *idisfer:inCrisisEvent* telle que *hasInstant*(e, i_1) [ou *hasPeriod*($e, [i_1; i_2]$)]
 - $egvr, egvrBegin$ et $egvrEnd$ des instances du concept *gvr:inCrisisEventGRV* telles que *represents*($egvr, e$) [ou *represents*($egvrBegin, e$) et *represents*($egvrEnd, e$)]
- \forall *hasInstant*(e, i_1) [ou *hasPeriod*($e, [i_1; i_2]$)] avec i_j (avec $j=1$ ou $j=2$) = (Y_j, M_j, D_j, H_j, Mn_j)
alors *hasInstant*($egvr, i'_1$) [ou (*hasInstant*($egvrBegin, i'_1$) et *hasInstant*($egvrEnd, i'_2$))] avec i'_j (avec $j=1$ ou $j=2$) = (Y_j, M_j, D_j)

4.3.4 Représentation du type d'événement

L'ontologie IDISFER classe les événements selon une hiérarchie de concepts dont la classe principale *genericEvent* est spécialisée en : *Phénomène naturel*, *Incident*, *Perturbation* et *Réaction* (voir sections 4.2.1. et 4.2.2). Permettre de distinguer sur la carte les différents types d'événement survenus pendant l'inondation est important pour favoriser la compréhension de la situation dans son ensemble. Le sous-concept d'événement dans l'ontologie est, du point de vue de sa restitution cartographique, assimilable à une donnée qualitative nominale. Selon les règles de la sémiologie graphique, la *forme* est une variable relativement adaptée pour représenter des données nominales et favorise l'association des objets semblables (Dibiase et al., 1992). Plus précisément, pour constituer la légende de la carte, l'emploi de symboles iconiques est intéressant (Slocum et al., 2001) car :

- en utilisant des métaphores connues des lecteurs de la carte, ils peuvent accélérer le décodage de l'information (la loi de familiarité de la *Gestalt*), sous réserve que leur forme reste simple pour favoriser la perception visuelle ;
- en contribuant à expliciter graphiquement la sémantique du concept qu'ils représentent, ils peuvent aussi favoriser la compréhension et la mémorisation de l'information.

Ces considérations amènent à préconiser, pour la restitution graphique des concepts d'une ontologie de domaine, le recours à des symboles évocateurs, aux formes stylisées, et, autant que possible, issus ou dérivés de chartes graphiques en vigueur.

De manière générale, dans une ontologie, le nombre de concepts, et leur structuration hiérarchique en différents niveaux peuvent être très variables. De plus, l'arbre des concepts n'est pas forcément équilibré, des branches pouvant être plus détaillées que d'autres. Par ailleurs, le nombre d'*instances* des concepts (et, par suite, de symboles) à afficher peut grandement différer d'un concept à l'autre, et justifier qu'une sous-classe, par exemple, ait son propre symbole.

En regard de ces considérations, il est également reconnu qu'une (catégorie de) légende doit idéalement être composée de 7 +/- 2 items pour rester lisible et mémorisable (Le Fur, 2007; Miller, 1956). Il paraît, dans ces conditions, difficile de tendre vers une règle systématique définissant les modalités d'association de symboles aux concepts d'une ontologie¹⁰⁷. De plus, selon le postulat à l'origine de notre approche, il peut être pertinent d'aller vers une autre organisation des concepts que celle utilisée dans l'ontologie d'origine, soit afin d'en donner une représentation plus simple, soit pour répondre aux besoins de visualisation des utilisateurs. Dans ce contexte, il est possible d'exploiter le graphe issu de l'ontologie initiale pour construire une autre classification des concepts dans l'ontologie *GVR Matching*, classification utilisée pour constituer en partie la légende.

Pour revenir à notre exemple, le choix d'attribuer un symbole à chaque catégorie d'événements telle que définie dans l'ontologie (*Phénomène naturel*, *Incident*, *Perturbation* et *Réaction*) n'est pas complètement satisfaisant en regard des usages de type REX attendus pour la géovisualisation. En ce qui concerne la restitution des incidents (classe *idisfer:inCrisisIncident*, une sous-classe de *idisfer:inCrisisEvent*), les catégoriser en fonction des classes d'objets ferroviaires impactées fait sens pour des experts métier ayant besoin d'identifier facilement les impacts des inondations sur leurs objets d'étude. Cela se traduit par la création, dans l'ontologie *GVR Matching*, d'un ensemble de classes obtenues par l'exploitation du graphe des instances d'événements dans IDISFER, en utilisant

¹⁰⁷ Par exemple, en nous basant sur la hiérarchie des concepts, décider d'associer un symbole à toutes les sous-classes des classes à un niveau n de profondeur, si elles sont en nombre inférieur à 7 (remonter d'un niveau sinon).

le prédicat *idisfer:impacts* qui relie un incident à un objet du système ferroviaire (voir Figure 26). Schématisée par la seule classe *grvm:incidentOnRSO*, il en existe en fait plusieurs, autant que de types d'objets *idisfer:railwaySystemObject* impactées par un incident dans la situation de crise étudiée (la chaîne de texte RSO est remplacée dans les 'vraies' classes). Les instances de concepts *grvm:inCrisisIncident* sont associées à ces nouvelles classes (prédicat *isClassifiedAs*). Chaque classe *grvm:incidentOnRSO* est par ailleurs une sous-classe du concept *geovis:LegendItem*, lui-même sous-concept de *geovis:GeoVisRepresentation*.

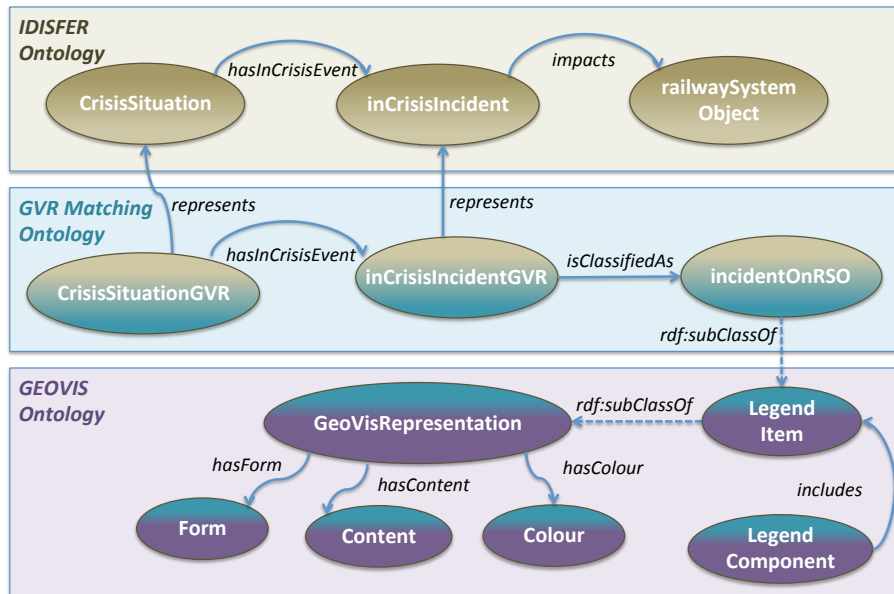


Figure 26 - Exemple de dérivation de concepts pour construire une classification exploitée dans la légende

On définit ainsi:

Soient

- s une instance du concept *idisfer:CrisisSituation*
 - et E l'ensemble des concepts de *idisfer:inCrisisIncident*
- tels que pour tout $e \in E$, *hasInCrisisEvent*(s, e)

Pour tout RSO^{108} tel que $\exists \text{ impacts}(e, RSO)$, avec *rdf:subClassOf*($RSO, \text{railwaySystemObject}$) alors

le concept *grvm:incidentOnRSO*¹⁰⁹ est défini dans l'ontologie (*)

et pour $egvr$ une instance du concept *grvm:inCrisisEventGRV* telle que *represents*($egvr, e$) alors *grvm:isClassifiedAs*($e, \text{incidentOnRSO}$) est ajouté

Dans l'instruction marquée par (*) présentée précédemment, la définition du concept *grvm:incidentOnRSO* est évoquée. Outre le nom donné à cette classe et la définition de ces propriétés, on note qu'en tant que sous-classe de *geovis:geovisRepresentation* une forme (predicat *hasForm*) lui est affectée. Il s'agit ici du symbole iconique évoqué en début de cette section. La Figure 27 (partie de gauche) donne un aperçu du résultat de la méthodologie que nous venons de décrire, et montre la symbologie attribuée aux incidents, re-catégorisés en fonction du type d'objet ferroviaire qu'ils ont endommagé (e.g. gare, objet de la maintenance, ou encore ouvrage d'art – OA, ou ouvrage hydraulique – OH). Les symboles employés sont issus de la charte graphique de l'entreprise SNCF.

Dans la catégorie des mesures de réaction (voir Figure 27, à droite), celles en lien avec des travaux se distinguent des autres car elles concernent les composants de l'infrastructure ferroviaire, qui sont les objets d'intérêt des utilisateurs cibles, tandis que les autres mesures de réaction concernent la régulation des circulations. Par conséquent, les travaux sont distingués visuellement des autres types

¹⁰⁸ RSO pour *Railway System Object*

¹⁰⁹ Formellement une concaténation de la chaîne « incidentOn » et du nom de la classe désignée par RSO est opérée.

de mesures de réaction par un symbole qui leur est propre. Dans ce cas à nouveau, la dérivation de cette représentation est obtenue à partir de l'ontologie originale en exploitant le graphe associé.

Des symboles iconiques ont été attribués aux événements de type incidents et aux phénomènes naturels, tandis que des symboles abstraits sont utilisés pour les perturbations des circulations et les mesures de réaction. Ce choix de deux types de visuel (iconiques *vs* abstraits) vise à augmenter les différences entre symboles afin de faciliter la catégorisation des événements lors de la lecture et leur intégration dans les représentations mentales (voir (Saint-Marc, 2017), Chapitre IV).

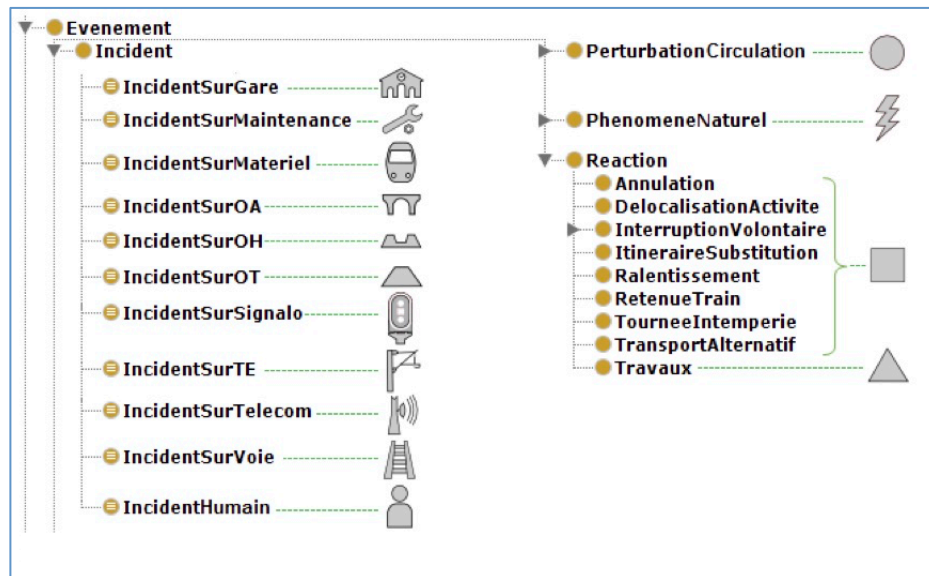


Figure 27 - Symboles affectés aux classes d'événements dérivées à partir de l'ontologie IDISFER

Les sous-classes d'Incident n'apparaissent pas telles quelles dans IDISFER mais ont été construites à partir du graphe pour mieux adapter la présentation des informations aux besoins de utilisateurs de la géovisualisation. Source : d'après (Saint-Marc, 2017).

4.3.5 Représentation de la date des événements

Selon la sémiologie graphique, la date peut être considérée soit comme une variable quantitative, soit comme une variable qualitative ordinale (voir (Saint-Marc, 2017), Chapitre V). La couleur est une variable visuelle adaptée à la fois à ces deux conceptions (Bertin, 1967). L'utilisation d'une échelle de valeurs progressive pour déterminer la couleur à appliquer selon la date des événements permet de distinguer les événements anciens des plus récents et, ainsi, de percevoir l'ordre chronologique.

Le grand nombre de dates à représenter sur la durée d'une situation pour en montrer l'évolution, nous incite à construire cette échelle de valeurs progressive en la structurant en différentes périodes, chacune associée à des teintes différentes. Cette approche reprend de précédents travaux¹¹⁰ que nous avons menés sur la représentation de série de données d'éruptions du Piton de la Fournaise (Davoine et al., 2012) entre 1970 et 2010. Une couleur était associée à une décennie (*e.g.* le vert pour les années 70). Le dégradé de couleur (*e.g.* d'un vert clair à un vert foncé) était utilisé pour rendre compte, à l'intérieur d'une même décennie, des différentes années lors desquelles une éruption avait eu lieu. L'approche a récemment fait l'objet, dans le cadre du projet ANR Eclats¹¹¹, du développement sous R d'une technique d'automatisation de création de palettes composées d'autant de couleurs et de niveaux de dégradés que nécessaire pour la constitution d'une légende (Watine, 2018).

¹¹⁰ Projets VOLCANISTE (financement: CNRS PEPS INS2I, 2012) puis ALEA-Volcanique (Financement: Ministère de l'Ecologie, de l'Energie, du Développement Durable et de la Mer, 2013) menés avec les partenaires suivants : Institut de Physique du Globe, Observatoire Volcanologique du Piton de la Fournaise, Observatoire Géosciences Réunion, Observatoire du Vésuve-Institut National de Géophysique et de Volcanologie.

¹¹¹ Projet Eclats : Extraction automatisée des Contenus géoLinguistiques d'Atlas et analyse Spatiale: application à la Dialectologie. 01/10/2015-30/09/2019. Financé par l'ANR sur l'appel Grands Défis Sociétaux et Défi de tous les savoirs - Le numérique au service des arts, du patrimoine, des industries culturelles et éditoriales – 2015. <https://eclats.imag.fr>

La représentation de la date des événements passe donc par l'attribution d'une valeur de couleur aux symboles proposés dans la section précédente (4.3.4). Cela requiert deux étapes : *i*) identifier à quelle période de la situation appartient un événement, et *ii*) lui attribuer une valeur de couleur selon son rang dans la période.

Segmentation de la situation en périodes. La nature du phénomène à l'origine de la situation observée possède une dynamique propre qui peut être un premier critère de décision pour segmenter la situation en différentes périodes. Ainsi, dans notre cas d'application, qui étudie l'impact des inondations sur le système ferroviaire, c'est ce phénomène qui va guider la segmentation. Le cycle de vie d'inondation est utilisé pour déterminer différents segments : la *période préalable à l'inondation* (au cours de laquelle ont parfois lieu des événements conduisant à l'inondation), la *phase de crue*, le *pic de crue* et la *décrue*. Différents incidents peuvent apparaître au cours de ces périodes.

Les périodes suivantes visent à structurer temporellement les réactions engagées, en donnant une vision des événements à *court*, *moyen* et *long termes*. Ces périodes sont segmentées automatiquement en détectant des discontinuités temporelles, c'est-à-dire des espacements entre deux événements successifs d'une durée relativement importante par rapport à l'espacement observé pour les événements précédents. On définit ainsi, en utilisant les concepts de l'ontologie *GVR Matching* (car les représentations d'événements y sont associées à des dates unitaires (voir section 4.3.3)) :

Soient

- *s* une instance du concept *grvm:CrisisSituationGVR* telle que

grvm:hasPeriod(s, p) avec *p* une instance du concept *grvm:CrisisSituationPeriodGVR*

par simplification, on dénote :

$$p \in \{\text{"Pré-Inondation"}, \text{"Crue"}, \text{"Pic de Crue"}, \text{"Décrue"}, \text{"Court-terme"}, \text{"Moyen-terme"}, \text{"Long-terme"}\}$$

- et $E = \{e_1, \dots, e_n\}$ l'ensemble des N instances e_i (avec $i=1\dots n$) de *grvm:inCrisisEventGVR*

tels que pour tout $e_i \in E$, $\exists \text{hasInCrisisEvent}(s, e_i)$

On pose $D = \{d_1, \dots, d_n\}$ est l'ensemble des dates d_i (avec $i=1\dots n$) associées aux e_i et

qui entretiennent une relation d'ordre chronologique ($d_i \leq d_{i-1}$)

On définit $P(e_i)$ la fonction associant une période à un événement en fonction de sa date d_i

telle que $P: e_i$ (avec $i=1\dots n$) $\rightarrow p$ i.e. création d'un prédicat *grvm: belongsTo*(e_1, p)

étant données :

d_j la date de e_j qui marque le début de l'inondation,

d_k la date de e_k qui marque le pic de la crue et

d_l la date de e_l qui marque la fin de la décrue

alors

$\forall d_i < d_j, P(e_i) = \text{"Pré-Inondation"}$

$\forall d_i$ tel que $d_j \leq d_i < d_k, P(e_i) = \text{"Crue"}$

$\forall d_i = d_k, P(e_i) = \text{"Pic de Crue"}$

$\forall d_i$ tel que $d_k < d_i \leq d_l, P(e_i) = \text{"Décrue"}$

$\forall d_i > d_l :$

$$\left\{ \begin{array}{l} \text{si } (d_i = d_{l+1}) \\ \text{alors } P(e_i) = \text{"Court-terme"} \\ \text{si } (d_i > d_{l+1} \text{ et } \text{grvm: belongsTo}(e_{i-1}, \text{"Court-terme"}) \text{ et } (d_i - d_{i-1}) >= \alpha (d_{i-1} - d_{i-2})) \\ \text{alors } P(e_i) = \text{"Moyen-terme"} \\ \text{si } (d_i > d_{l+1} \text{ et } \text{grvm: belongsTo}(e_{i-1}, \text{"Moyen-terme"}) \text{ et } (d_i - d_{i-1}) >= \alpha (d_{i-1} - d_{i-2})) \\ \text{alors } P(e_i) = \text{"Long-terme"} \\ \text{si } (d_i > d_{l+1} \text{ et } \text{grvm: belongsTo}(e_{i-1}, \text{"Long-terme"})) \\ \text{alors } P(e_i) = \text{"Long-terme"} \end{array} \right.$$

où α est un facteur seuil au-delà duquel un intervalle temporel est considéré comme suffisant pour justifier un changement de période¹¹².

¹¹² Pour les cinq cas d'étude d'inondations traités dans la thèse de C. Saint-Marc, il est proposé $\alpha = 6$ (i.e. si un intervalle entre deux événements est au moins 6 fois plus grand que l'intervalle séparant les deux derniers événements, alors il faut changer de période). Ce chiffre et le mode de calcul de détection du changement de période, nécessiteraient cependant d'être validés sur un plus grand nombre de cas d'études et pour des phénomènes de natures différentes à des fins de généralisation.

À ce stade, chaque événement est donc associé à une période de la situation observée. L'étape suivante est de déterminer les couleurs à utiliser pour distinguer visuellement les périodes les unes des autres, mais aussi la chronologie des événements à l'intérieur de certaines d'entre elles.

Attribution d'une couleur à chaque date. Le même niveau de précision n'est pas attendu dans la distinction visuelle (au sens de la perception) des dates sur l'ensemble de la situation : dans les premiers jours de la situation, il est utile de distinguer les dates précises des événements (phénomènes et incidents) qui se succèdent, pour reconstituer leur chronologie. En revanche, percevoir précisément les dates unitaires de tous les événements de travaux est inutile. La perception de leur ordre et de leur répétition au cours du temps a été jugée suffisante pour appréhender le récit de la situation sur ce point (voir (Saint-Marc, 2017), Chapitre VI). Cela, associé aux caractéristiques de quantité et d'étendue temporelle des événements à restituer, a des conséquences différentes sur les choix de représentation des dates selon les périodes.

Les événements de la période préalable à l'inondation sont généralement peu nombreux et sont représentés dans une couleur neutre (un ton crème) choisi car discriminant par rapport à l'ensemble des autres couleurs employées. Les périodes relatives à l'inondation en elle-même et les phases de crue et de décrue sont, en revanche, marquées par des événements nombreux, sur des dates elles-aussi nombreuses et rapprochées. L'affectation de couleurs selon une double progression rouge/vert est utilisée pour différencier les deux phases. Les couleurs sont choisies en fonction de leur symbolique dans la culture occidentale : le rouge associé au danger pour la phase de crue et le vert pour traduire le retour au calme de la phase de décrue. Le pic de crue est symbolisé en jaune, couleur entre le vert et le rouge sur un cercle chromatique tel que celui de Ewald Hering (voir Figure 28). Les couleurs utilisées sont saturées et intenses pour rendre saillants les événements survenus dans cette période car ils sont *a priori* les plus importants pour le public cible.

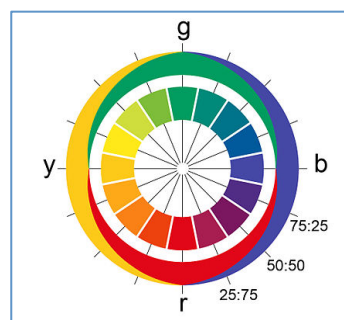


Figure 28 - Reproduction du Cercle chromatique Ewald Hering.

source : https://fr.wikipedia.org/wiki/Ewald_Hering

Les événements à court-terme sont représentés par une échelle déclinant la couleur bleue, couleur voisine du vert – celle utilisée pour représenter la période précédente. L'objectif est de favoriser la perception d'une continuité dans le déroulement temporel, malgré le changement de teinte. Un dégradé est utilisé pour marquer chaque date distinctement dans cette période, car, là-encore, les événements sont nombreux à être observés à des dates elles-aussi nombreuses et rapprochées.

Les événements à moyen-terme et à long-terme, qui sont plus rares et très espacés dans le temps, ont été regroupés sous une teinte unique, sans dégradé de couleur. Il paraît en effet moins utile de distinguer visuellement les dates précises de ces événements, dates qui sont, par ailleurs, rarement citées à la granularité du jour dans les documents. Suivant le cercle chromatique, c'est une couleur entre le bleu et le rouge qui est retenue.

Cette approche, qui allie couleurs et dégradés de celles-ci et exploite la continuité du cercle chromatique, vise à favoriser la perception des différentes catégories d'informations. Une méthode définie dans (Saint-Marc, 2017) associe une couleur à chaque période. Lorsqu'un dégradé de couleur est requis (*i.e.* pour les périodes de crue, de décrue et de réactions à court-terme), les variations de la teinte sont choisies pour maximiser leur différenciation visuelle sur la plage de valeurs. Dans le cadre de notre approche, cette méthode peut être implémentée pour affecter la couleur souhaitée, via le prédicat `grvm:hasColour`, à chaque instance d'un concept `grvm:DateInLegend` qui est une sous-classe

de *geovis:LegendItem*. Chaque instance de *gvrvm:inCrisisEventGVR* est associée via le prédicat *gvrvm:correspondsTo* à l'instance de *gvrvm:DateInLegend* à laquelle elle correspond étant donnée sa date¹¹³.

La Figure 29 donne un exemple d'un item de légende instancié à partir du concept *gvrvm:DateInLegend*. Il est composé d'une forme rectangulaire, du contenu « 29/12/1947 » et de la couleur *RVB(255,100,104)* (variation d'un rouge, couleur affectée à la période de crue). Tous les événements *gvrvm:inCrisisEventGVR* associés à cette instance sont datés du « 29/12/1947 ».

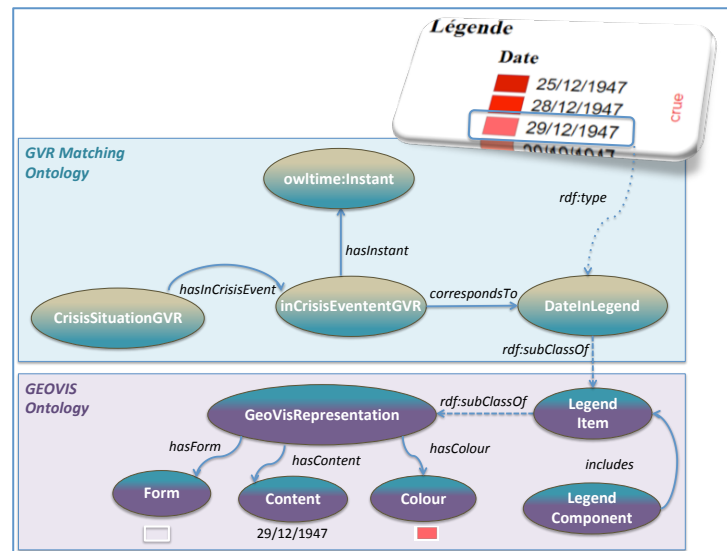


Figure 29 - Illustration de l'instanciation du concept *DateInLegend* de l'ontologie *GVR Matching*.

La Figure 30 montre un exemple complet de légende construite avec cette approche. Deux modes de visualisation des dates sont implémentés dans la légende : un mode dit *contracté* (ou *plié*) et un mode *dilaté* (ou *déplié*). Le mode contracté insiste sur la chronologie générale de la situation, plus adapté à l'appréhension des périodes dans leur globalité, tandis que le mode dilaté met l'accent sur les dates unitaires, qu'il est plus important d'identifier dans les phases d'inondation et de réactions à court-terme. Ces deux modalités mettent donc en valeur des temporalités différentes des événements lors de la lecture de la carte.

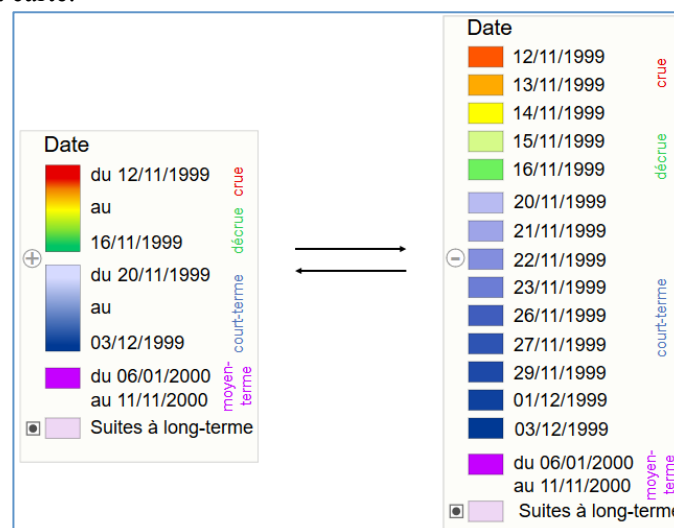


Figure 30 - Exemples de légendes rendant compte des périodes de la situation de crise

Modes contracté (à gauche) et dilaté (à droite) de la légende. La légende contractée est affichée par défaut dans l'interface du prototype développé et l'utilisateur peut choisir interactivement de passer à la légende dilatée en fonction de sa tâche. Source : (Saint-Marc, 2017).

¹¹³ Dans la suite, nous ne donnons plus les expressions semi-formelles correspondantes pour alléger le propos.

La dérivation des concepts de l'ontologie GRV Matching décrite dans le paragraphe précédent correspond au cas de la légende en mode déplié, car elle repose sur la création d'une instance de *grvm:DateInLegend* par date à afficher. Pour construire la légende en mode contracté à partir de l'ontologie, la création de concepts et d'instance dédiés est nécessaire (puisque un item de légende regroupe dans ce cas plusieurs dates et que la couleur associée doit rendre compte du dégradé).

4.3.6 Visualisation d'événements de même localisation

Plusieurs événements successifs peuvent être survenus au même emplacement au cours du temps. Dans l'exemple présenté dans la Figure 23, un phénomène naturel (embâcle), deux dommages (chute de la pile du pont et déformation du tablier), et un événement de type travaux se succèdent sur le même ouvrage ferroviaire. Des problèmes de densité et de superposition visuelle entre les symboles sur la carte sont donc à traiter.

Des méthodes de gestion d'entités superposées proposées dans la littérature (voir (Saint-Marc, 2017), chapitre II) peuvent être envisagées. Comme nous souhaitons que les types d'événement et leurs dates unitaires restent visibles malgré la densité de points, nous retenons la méthode de déplacement des points par rapport à leur position initiale. Nous nous inspirons de la carte publiée par Tufte (1990) et proposons de construire des « piles chronologiques » d'événements, illustrées dans la Figure 31. Le principe consiste à :

- éloigner chaque symbole correspondant à un événement de sa localisation initiale, c'est-à-dire de la localisation de l'objet impacté. Dans la Figure 31, les objets impactés sont des point kilométriques de la voie ferrée, ou des ouvrages sur celle-ci.
- à aligner verticalement l'ensemble des symboles d'événements de même localisation, ordonnés chronologiquement ;
- à relier cette construction à la localisation concernée par une ligne.

L'illustration (Figure 31) montre que les symboles empilés sont placés dans un cadre blanc qui vise à les abstraire du fond de carte, tout en suggérant qu'ils ne sont pas situés réellement à la place qu'ils occupent sur la carte (*i.e.* la portion recouverte par la pile). Le cadre blanc est relié par une ligne à leur localisation véritable (*i.e.* celle de l'objet concernés par les événements).

Lorsque le nombre d'événements survenus en un même lieu est trop grand pour qu'un cadre blanc soit visible dans son intégralité dans la carte, ce cadre est segmenté en plusieurs cadres, placés de part et d'autre du symbole correspondant à l'objet impacté, en fonction de l'espace graphique disponible.

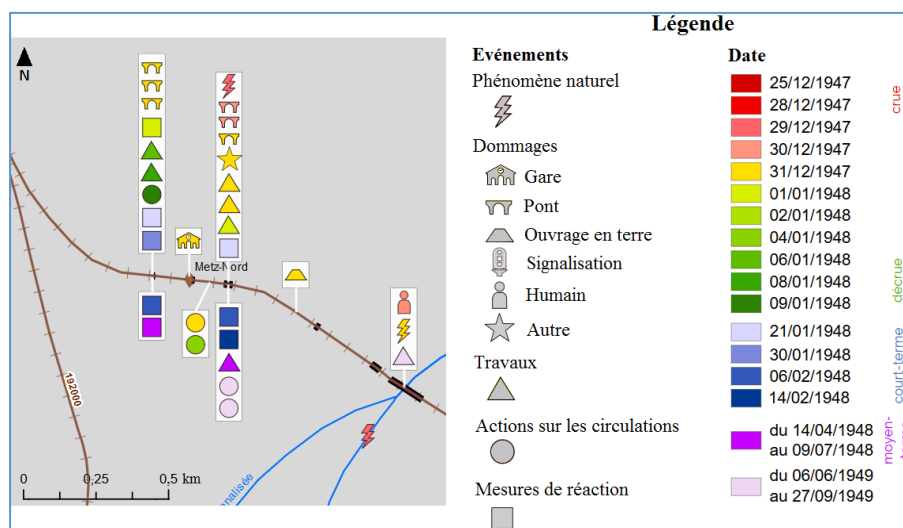


Figure 31 - Illustration de la méthode de la pile chronologique

Les événements survenus à un même endroit au cours du temps sont ordonnés chronologiquement dans un cadre (le plus ancien de la série est placé en haut). Source : image extraite de (Saint-Marc, 2017).

Nous renvoyons le lecteur à (Saint-Marc, 2017), Chapitre VI, pages 235-236, pour la description d'une première proposition de méthode¹¹⁴ qui détermine le positionnement des cadres dans la carte. Dans notre approche, le concept de pile chronologique (*grvm:ChronoStack*) est ajouté dans l'ontologie *GRV Matching*, comme sous-classe de la classe *geovis:GeoVisRepresentation* et de *rdf:List* (car l'ordre des éléments a de l'importance). Toutes les instances d'événements (*grvm:InCrisisEventGVR*) qui impactent le même objet (et, par conséquent, sont à localiser au même endroit) sont ainsi associées à une instance de pile chronologique (on utilise à cet effet la valeur du prédicat *grvm:hasGeometry*). L'ordre des éléments est déterminé par leurs dates (prédicat *grvm:hasInstant*), considérées dans l'ordre croissant¹¹⁵.

4.3.7 Visualisation des relations de causalité entre événements

L'évolution d'une situation de crise, dans notre contexte, est intrinsèquement portée par les relations de cause à effet qu'entretiennent les différents événements. Figurer explicitement dans la géovisualisation les effets dominos est identifié comme pertinent dans la cadre d'un REX (voir (Saint-Marc, 2017), Chapitre I). Par ailleurs, la perception des relations de causalité entre événements est susceptible d'aider à la compréhension et à la mémorisation d'un récit (voir (Saint-Marc, 2017), Chapitre V). Les relations de causalité entre événements vont-être représentées par des lignes courbes pleines reliant chaque événement de type « cause » à sa ou ses « conséquences » selon une méthode de restitution dans la carte exposée dans (Saint-Marc, 2017), Chapitre VI, pages 236-237, qui détermine notamment les positions de départ et d'arrivée de la courbe, ainsi que son incurvation.

Les lignes de causalités étant nombreuses, les afficher toutes simultanément surchargerait la carte et masquerait les symboles d'événements. Ces lignes ne sont donc affichées qu'à la demande de l'utilisateur : en cliquant sur le symbole d'un événement, les lignes le reliant à sa (ses) cause(s) et à sa (ses) conséquence(s) apparaissent (voir Figure 32). La couleur des lignes distingue les causes, reliées par des lignes rouges à l'événement cliqué, et les conséquences qui sont liées à l'événement cliqué par des lignes noires¹¹⁶.

L'ensemble de ces considérations est traduit dans notre approche par l'ajout de différents concepts à l'ontologie *GVR Matching*, qui seront, une fois encore, instanciés par exploitation des données issues de l'ontologie *IDISFER* ou de celles déjà présentes dans *GVR Matching*. On note ainsi :

- la création de *grvm:CauseLegendItem* et *grvm:ConsequenceLegendItem*, deux concepts sous-classes de *grvm:LegendItem*, associés notamment aux couleurs (*gvrvm:hasColour*) rouge et noire respectivement et à des lignes courbes (*gvrvm:hasForm*).
- la création des concepts *grvm:listOfCauses* et *grvm:listOfConsequences* auxquels est associé un *grvm:InCrisisEventGVR* par les prédicats *grvm:hasCauses* et *grvm:hasConsequences*.
- Le concept *grvm:listOfCauses* est lié au concept *grvm:CauseLegendItem* pour indiquer les conventions graphiques à exploiter pour représenter les causes. De même, le concept *grvm:listOfConsequences* est lié au concept *grvm:ConsequenceLegendItem* pour représenter les conséquences.
- pour chaque événement *e* de *grvm:InCrisisEventGVR*, on procède à l'instanciation des concepts *grvm:listOfCauses* et *grvm:listOfConsequences* grâce au parcours du graphe de l'ontologie *IDISFER* (voir Figure 22). La liste des causes de *e* est donnée par les événements *c* sujets des prédicats *idisfer:causes(c, e)*. La liste des conséquences de *e* est donnée par les événements *c* objets des prédicats *idisfer:causes(e, c)*.

¹¹⁴ Cette méthode tient compte notamment du nombre d'événements à afficher pour décider de scinder la pile en plusieurs cadres, mais aussi de la présence d'autres cadres à proximité de celui à afficher pour optimiser le positionnement de l'ensemble des piles chronologiques.

¹¹⁵ La notion de pile est à prendre ici au sens visuel du terme : les symboles sont empilés. L'ordre des éléments dans la pile, rangés en ordre croissant des dates, est l'inverse de la visualisation puisqu'en bas de la pile sont visibles les éléments les plus récents. Cela est motivé par les effets d'animation décrits en section 4.4.

¹¹⁶ Notons que dans le cadre de cette recherche, nous avons testé la variable couleur auprès des utilisateurs, mais d'autres méthodes pourraient être employées (par exemple des variations de texture : trait plein, pointillé).

- les symboles (*geovis:Form*) associés aux représentations d'événements *grvm:InCrisisEventGVR* deviennent des sous-classes de *geovis:InteractionComponent* pour être dotés de capacités à réagir aux actions de l'utilisateur.

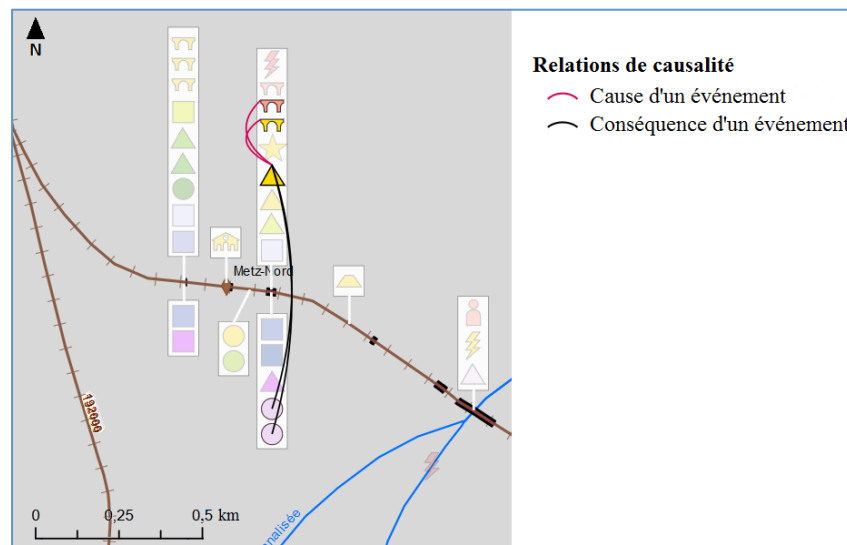


Figure 32 - Affichage à la demande des relations de causes et conséquences d'un événement.

À droite : la légende. À gauche : visualisation du résultat suite un clic sur le symbole d'un événement (ici des travaux). Les autres événements non liés par ces relations et le cadre des piles chronologiques s'estompent pour favoriser la perception des symboles de l'événement sélectionné, de ses causes (lignes rouges et symboles reliés) et de ses conséquences (lignes noires et symboles reliés). Source : image extraite de (Saint-Marc, 2017).

4.3.8 Visualisation des relations de causalités lointaines

Les événements ont une répartition spatiale hétérogène. S'ils se concentrent souvent dans certaines zones (en majorité le long des axes ferroviaires, plus vulnérables aux phénomènes hydro-climatiques), certaines situations de crise liées à des inondations majeures peuvent impacter de larges zones, à l'échelle d'une région. La restitution visuelle de ces situations risque d'aboutir à de larges espaces vides, à petite échelle, alternant avec des zones où les événements sont trop proches pour être distingués les uns des autres. Au delà de doter l'interface des fonctionnalités de zoom et de déplacement dans l'emprise de la carte, une solution proposée dans (Saint-Marc, 2017) vise à maintenir une vue globale sur la situation nécessaire pour la perception des effets dominos. Elle répond au besoin de représenter les relations existantes entre des événements spatialement distants qui ne figureraient pas dans l'emprise de la carte en cours de consultation. Un « encart contextuel » est ainsi associé à la carte, dans lequel s'affichent tous les événements qui sont en relation avec des événements situés dans l'emprise de la carte principale, mais qui se trouvent hors de cette emprise (voir Figure 33). L'encart donne une vue globale de la zone touchée par l'inondation et indique que des événements sont situés quelque part dans cet espace. Ce procédé visuel correspond à l'affichage d'une partie du contexte spatial en parallèle de la vue principale, selon la méthode du « *focus + context* » mais peut également être rapproché de la méthode « *overview + detail* » (voir (Saint-Marc, 2017), Chapitre II), puisque dans la carte l'encart permet de localiser la zone détaillée (voir le point bleu qui situe la ville de Metz, sur laquelle est centrée la description de la situation).

Dans (Saint-Marc, 2017), Chapitre VI, pages 239-240, une méthode est proposée pour déterminer s'il convient de déclencher la création de l'encart et d'en calculer les paramètres. Pour notre approche ontologique, des procédés similaires à ceux décrits dans les sections précédentes sont employés pour la création du concept correspondant à l'encart – et de ses relations avec les autres concepts – dans l'ontologie *GRV Matching*. Une des étapes consiste à déterminer les événements dont la représentation est à déporter dans l'encart. Pour cela, pour un événement *e* de *grvm:InCrisisEventGVR* sur lequel on clique, on parcourt la liste des instances d'événements répertoriées dans *grvm:listOfCauses* et *grvm:listOfConsequences* et à tester leur inclusion spatiale dans les limites de la carte détaillée (par le biais d'une requête *geosparql* sur les géométries des objets représentés, *i.e.* symbole et carte).

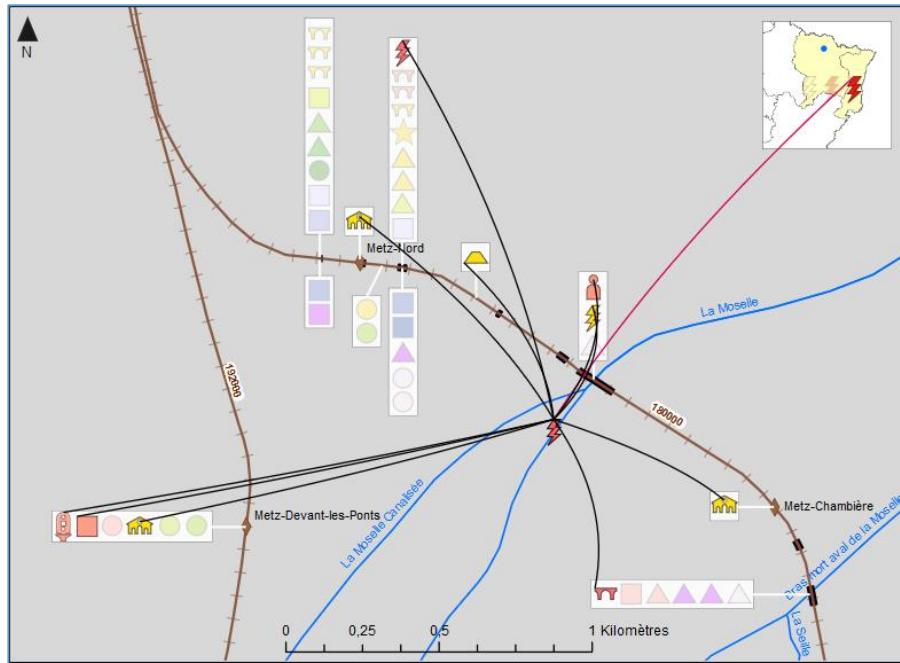


Figure 33 - Affichage d'une relation de causalité distante.

L'événement (un phénomène naturel symbolisé par un éclair) au centre de la carte principale résulte d'un événement spatialement distant (également symbolisé par un éclair) affiché dans l'encart contextuel accompagnant la carte. Source : image extraite de (Saint-Marc, 2017).

4.3.9 Visualisation des temporalités

Une vue de la chronologie de l'ensemble de la situation peut favoriser la compréhension de celle-ci. Dans ce but, une ligne de temps est proposée, avec comme pas de temps le jour conformément à la granularité retenue comme dénominateur commun pour les événements dans notre modèle (voir section 4.3.3). Les événements sont distribués par date sur un axe horizontal représentant le temps. L'axe vertical du graphique est employé pour montrer un attribut associé aux événements, par exemple leur type dans la Figure 34.

La ligne de temps reprend la symbologie des événements utilisée dans les autres composants : la forme du symbole pour représenter le type d'événement (voir section 4.3.4) et la couleur pour la date (voir section 4.3.5). La correspondance des propriétés visuelles (couleur et forme) associées aux événements entre les deux vues permet une meilleure mise en relation des informations figurant dans la ligne de temps avec celles figurant sur la carte.

La construction d'une représentation rendant compte de la chronologie de la situation nécessite de pallier deux problèmes :

- Premièrement, les événements sont survenus sur de longues périodes de temps et sont concentrés en majorité dans les premières semaines de l'inondation, laissant de longues périodes vides dans le graphique. Pour composer avec ces espacements temporels, la solution adoptée est de condenser graphiquement les périodes vides les plus longues, sous la forme d'un seul intervalle, auquel est attribuée une sémiologie différente qui contraste avec le reste de la ligne de temps : ainsi, dans la Figure 34, un arrière-plan de couleur grise montre les sauts temporels. Les événements à moyen et long-terme survenus entre ces périodes vides, qui sont peu nombreux et eux-mêmes épars dans le temps, sont également groupés dans un même intervalle temporel pour lequel les dates de début et de fin sont données.
- Deuxièmement, les événements qui sont survenus le même jour en début d'inondation apparaissent graphiquement superposés dans la ligne de temps (voir, par exemple, le 31/12/1947, les événements en jaune dans la Figure 34). Pour répondre à cet inconvénient résultant d'une graduation régulière de la ligne de temps (*i.e.* chaque jour occupe le même espace graphique), une seconde représentation du temps, proportionnel au nombre d'événements, est proposée. Dans cette réalisation, la largeur d'une journée est proportionnelle au nombre d'événements survenus ce jour (voir Figure 35). Nous nommons ce procédé le « déploiement de la ligne de temps ».

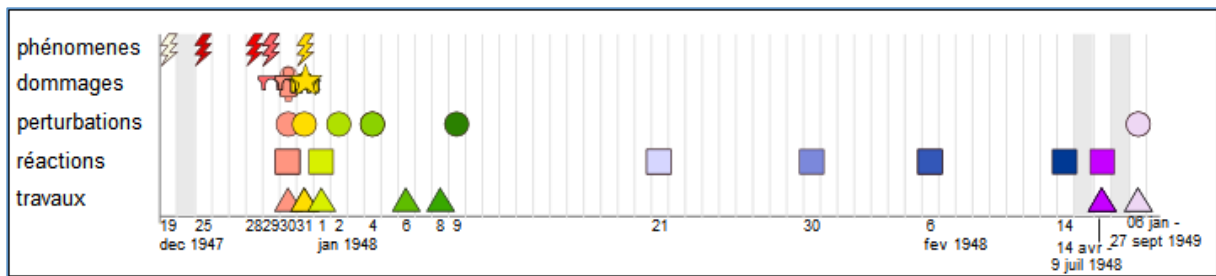


Figure 34 - Ligne à temps régulier montrant les événements survenus suite à l'inondation de 1947 à Metz.

La graduation de l'axe horizontal est celle du jour, excepté i) pour les tranches grises qui couvrent plusieurs jours correspondant à des périodes sans événement, et ii) pour les périodes en fin de ligne dédiées aux événements à moyen et long termes qui traduisent des périodes longues (du 14/04/1948 au 09/07/1948, puis du 06/01/1949 au 27/09/1949) sur lesquelles sont observés les événements.

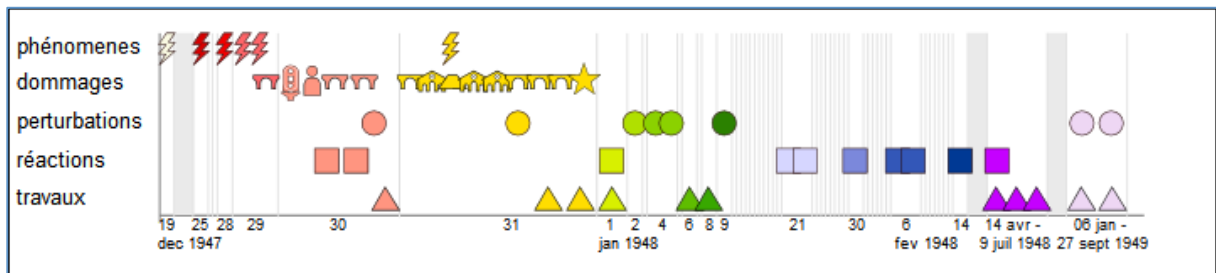


Figure 35 - Ligne à temps proportionnel montrant les événements survenus suite à l'inondation de 1947 à Metz.

La graduation de l'axe horizontal est celle du jour comme précédemment mais la largeur d'un jour est proportionnelle au nombre d'événements survenus à cette date. Les exceptions de représentation portant sur les périodes sans événement et périodes à moyen et court-termes sont celles évoquées pour la Figure 34.

Les deux lignes de temps¹¹⁷ mettent en valeur des temporalités différentes, parmi les types de temporalités identifiés dans la littérature (voir (Saint-Marc, 2017), Chapitre V, section 1.3) :

- La ligne de temps régulière permet de visualiser la date d'un événement et l'espacement temporel entre événements, et donne une idée de l'étendue des périodes structurant la situation (par exemple, par le jeu des couleurs, on perçoit que la décrue a lieu du 1^{er} au 9 janvier 1947. L'association d'une dimension attributaire en ordonnée nous semble également permettre indirectement de déduire des informations sur la fréquence des événements d'un même type.
- La ligne de temps proportionnel aux événements permet de visualiser la date, le taux de changement (*i.e.* la quantité de modifications survenues dans l'espace par unité de temps) et de mieux distinguer les types des événements. En revanche, il convient de noter que les événements survenus un même jour sont classés dans un ordre logique d'apparition (défini à partir des causalités), mais rien ne précise ne le fait qu'ils soient arrivés quasi-simultanément (dans un intervalle de quelques secondes) ou sont espacés de dizaines d'heures. La ligne de temps proportionnel aux événements donne donc une vision ordinale de la date et non numérique.

En vue de la construction de ces composants à intégrer dans une géovisualisation, il est possible de s'appuyer sur des concepts déjà définis dans les sections précédentes pour l'ontologie *GVR Matching*. Par exemple, en section 4.3.5, nous avons défini le concept *gvr:DateInLegend* et le prédicat *gvr:correspondsTo* qui permet d'obtenir la liste des événements *gvr:InCrisisEvent* associés à un jour particulier. Les conventions graphiques de couleur et de forme de symboles à utiliser dans les lignes de temps sont données respectivement par *gvr:DateInLegend* et *gvr:indincidentOnRSO* (voir section 4.3.4).

Cette réutilisabilité des concepts dans la construction de différentes formes de restitution visuelle d'une situation est l'un des intérêts de l'approche que nous avons introduite ici et qui repose sur le recours à une ontologie de *matching* entre les concepts d'une ontologie d'application, métier et/ou de domaine, et ceux d'une ontologie générique décrivant une géovisualisation.

¹¹⁷ Une troisième ligne de temps, plus classique car restituant le nombre d'événements par jour, existe également. Elle est visible dans la Figure 36.

Dans la section suivante, nous montrons comment les différents composants présentés jusqu'ici sont intégrés dans un environnement de géovisualisation, interactif et animé, permettant à l'utilisateur d'explorer les liens entre dimensions spatiales et temporelles et d'accéder à plus d'informations.

4.4 Prototype de géovisualisation dynamique du récit des événements

4.4.1 Vue d'ensemble

Un prototype d'environnement de géovisualisation, organisant dans une interface cohérente les propositions exposées précédemment a été développé¹¹⁸. Les langages HTML/CSS sont utilisés pour la mise en page, SVG/SMIL pour les vues graphiques et l'animation, et Javascript pour l'interactivité. Pour montrer la faisabilité de la mise en œuvre, les fichiers de jeux de données test ont été produits au format JSON, à partir de l'ontologie IDISFER¹¹⁹.

Les Figure 36 et Figure 37 illustrent deux vues de ce prototype, dont l'interface intègre les composants suivants :

- Une carte centrale, dans laquelle on peut voir les *pires chronologiques* (voir section 4.3.6) complètes (Figure 36) ou en cours de constitution (Figure 37), selon un procédé que nous décrivons plus loin.
- Une ligne de temps (située sous la carte), proposée en version dite « *synthétique* » (voir (Saint-Marc, 2017) page 249) qui donne un histogramme du nombre d'évènements survenus à chaque date (Figure 36) ou sous la forme de la ligne de temps proportionnel au nombre d'évènements, décrite en section 4.3.9, pour la Figure 37.
- Une fenêtre descriptive (sur la gauche) donnant accès aux descriptions détaillées et sous forme rédigée des événements (voir Figure 36, onglet *Détail des e^{vs}*). Une autre option permet d'obtenir une synthèse des événements par jour (voir Figure 37, onglet *Synthèse par jour*). Dans les deux cas, la case à cocher intitulée *Conserver l'historique* permet à l'utilisateur de compiler les descriptions des événements cliqués selon l'ordre de leur sélection. Par exemple, la Figure 37 illustre que l'utilisateur a consulté successivement les jours 11, 12 et 13 de la situation. Cela lui permet de retracer la chronologie des faits qui l'intéressent plus particulièrement dans sa démarche d'exploration.
- Une légende (sur la droite) qui est notamment composée des symboles décrivant le type des événements, leur datation et les relations de causalités (décrits respectivement en sections 4.3.4, 4.3.5 et 4.3.7).
- des menus déroulants pour modifier l'apparence de la carte ou de la ligne de temps (situés en haut) et un panneau de contrôle de l'animation complète cette interface.

L'ensemble de ces composants inclut des éléments interactifs dont l'un des effets est de synchroniser les différentes zones de l'interface : toute action effectuée sur l'un des éléments entraîne la mise à jour des contenus dans les autres zones et/ou leur mise en évidence par des procédés graphiques adaptés. En cela, l'outil reprend des principes éprouvés de la littérature que nous avons par ailleurs déjà exploités dans différents outils développés au sein de l'équipe, dans le domaine des risques naturels (voir une synthèse dans (Davoiné, 2014)) ou encore des données socio-économiques (Villanova-Oliver et al., 2013).

Cette interface intègre des fonctions classiques (Zanin et al., 2013) caractérisant les environnements de géovisualisation : le zoom et le déplacement dans la carte, la sélection de contenu par clic, le filtrage des contenus, le choix du fond de carte parmi une liste de choix (comme, par exemple, une photo aérienne dans la Figure 36 ou un fond simplifié et neutre dans la Figure 37).

Dans l'état de l'art qu'elle a réalisé, C. Saint-Marc souligne que de nombreuses géovisualisations dédiées au récit intègrent des cartes animées (cf. (Saint-Marc, 2017), Chapitre V, section 4.3.), notamment car la simple localisation des événements ne suffit pas à donner du sens au récit, alors que leur mise en séquence révèle toute leur signification spatiale (Caquard and Cartwright, 2014). Nous pensons également que l'animation est un vecteur favorisant la perception des effets dominos, et donc

¹¹⁸ Une vidéo de présentation du prototype est téléchargeable à <http://membres-lig.imag.fr/saintmar/html/bonus.html>

¹¹⁹ Nous rappelons que les travaux présentés précédemment et qui reposent sur le recours aux ontologies *GVR Matching* et *GEOVIS* sont postérieurs à la thèse de C. Saint-Marc, cadre dans lequel a été développé le prototype.

donnent une idée plus proche de la réalité de l'évolution d'une situation dans toute sa durée. Pour tester cette hypothèse, nous avons proposé une version animée de la carte intégrée dans le prototype de géovisualisation.

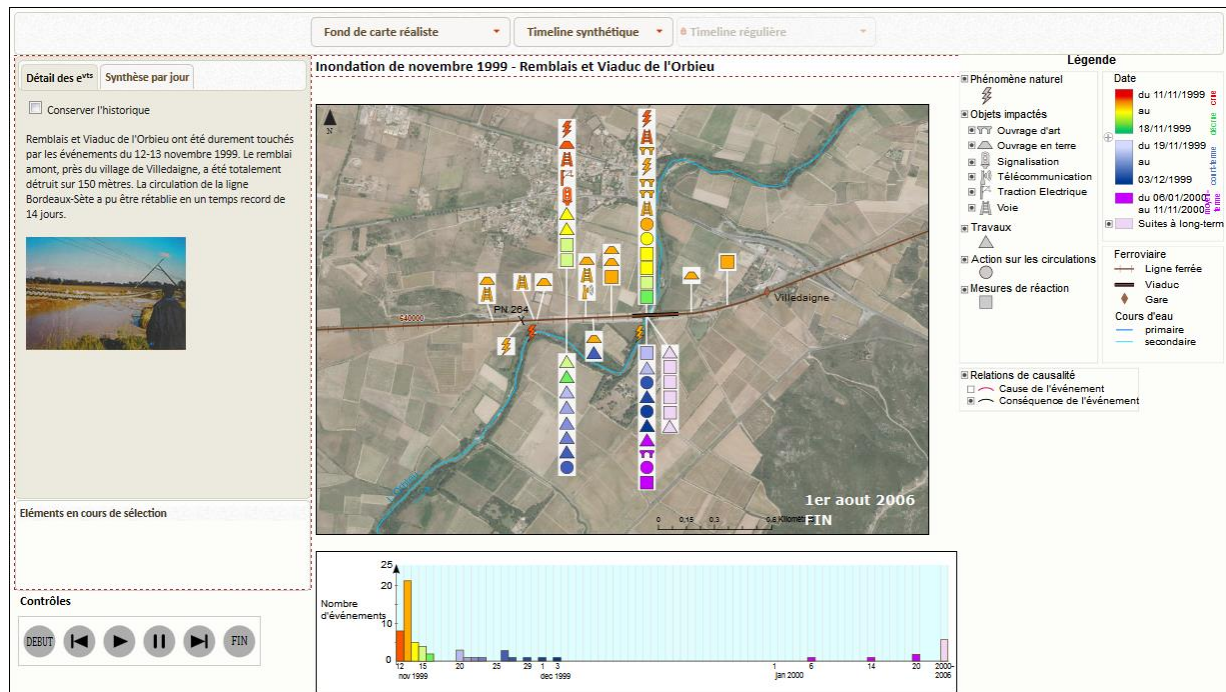


Figure 36 - Prototype de géovisualisation dédié au récit d'une situation de crise – Vue 1

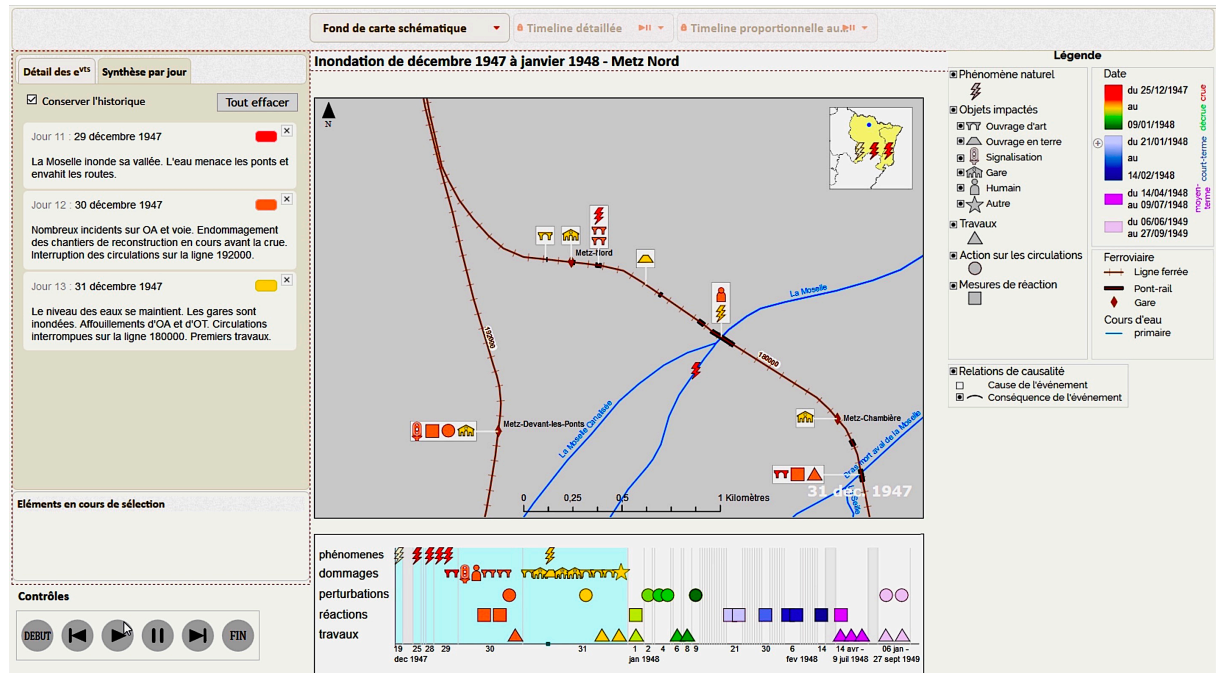


Figure 37 - Prototype de géovisualisation dédié au récit d'une situation de crise – Vue 2

4.4.2 Animation du récit de situation de crise

Les propositions de représentation des événements et de leurs relations de causalité ont été converties au format animé. Dans cette visualisation, l'animation suit un « scénario préconstruit » qui ne peut pas être modifié par l'utilisateur, selon la typologie de (Klein, 2007). L'animation est en revanche dotée des contrôles interactifs classiquement offerts.

L'apparition des événements est répartie dans des scènes successives, ordonnées selon l'ordre chronologique de l'histoire de l'inondation. Un événement et sa (ses) conséquence(s) apparaissent dans deux scènes distinctes, séparées par une scène d'animation rendant compte du lien de causalité. Le temps s'écoulant dans l'animation n'est pas proportionnel au temps réel.

La Figure 38 illustre l'enchaînement de scènes de l'animation¹²⁰. Les images sont à lire selon le sens indiqué par les flèches ajoutées en surimpression, soit de gauche à droite et de haut en bas. Elles constituent un extrait de l'animation associée au cas de l'inondation présenté dans la Figure 37. La 1^{ère} image montre l'événement symbolisé par un éclair correspondant à l'inondation du 29/12/1947, lorsque la rivière Moselle inonde la vallée, et ses conséquences observées le 31/12/1947. Les piles chronologiques déjà présentes dans la 1^{ère} image sont celles qui ont été construites auparavant lors de l'animation. Cinq lignes noires partant de ce symbole vont en direction des localisations d'objets qui ont subi les conséquences directes de cette inondation le 31/12/1947. On observe également en haut de l'image une autre ligne noire, partant d'un autre symbole éclair, en direction d'un point sur la ligne ferroviaire. Il s'agit d'un incident induit par le premier événement (car il est dans une pile chronologique) qui provoque lui même d'autres dommages. Ces lignes noires sont parcourues lors de l'animation par un figuré circulaire jaune qui montre l'orientation de la relation, de la cause vers la conséquence. La 2^{ème} image illustre comment, à la suite de l'animation de causalité, les symboles des incidents conséquences apparaissent en bout de ligne, c'est-à-dire à la localisation des objets impactés. Les symboles rejoignent ensuite la pile chronologique à laquelle ils sont associés (3^{ème} image). Le résultat de cette séquence de l'animation est illustré par la 4^{ème} image.

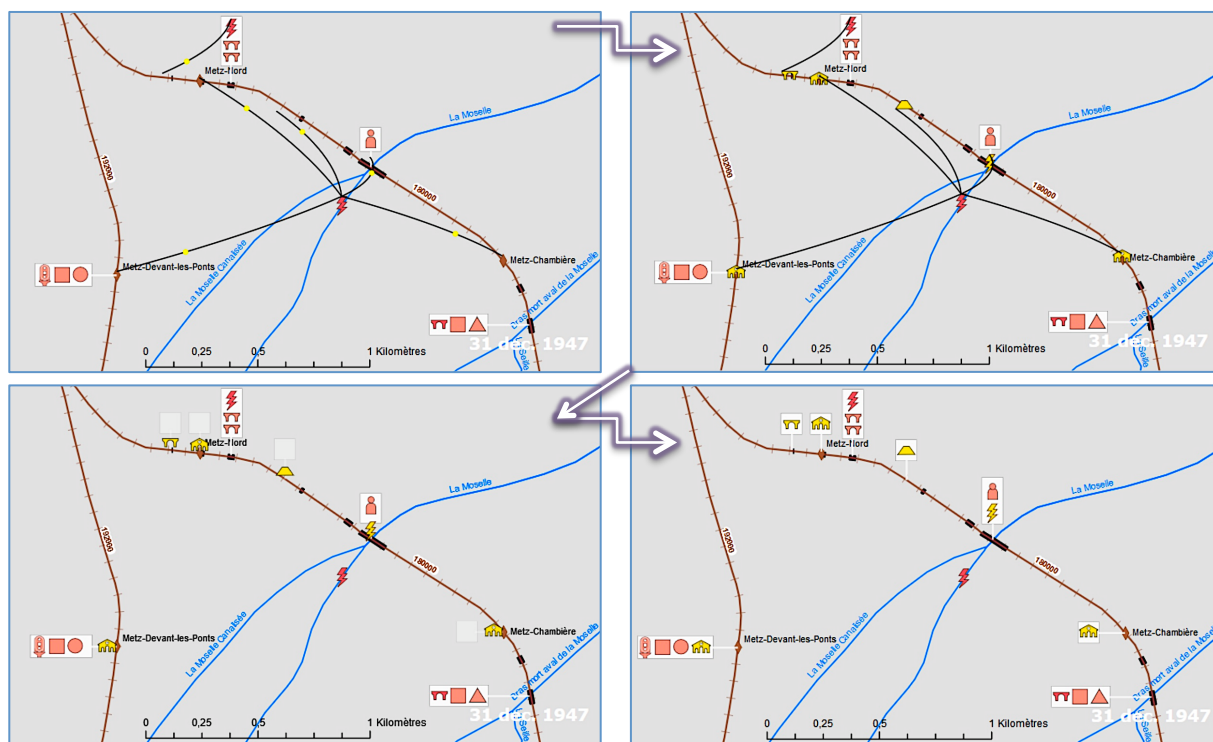


Figure 38 - Illustration de l'animation de la carte pour rendre compte des relations entre événements

Dans cette animation, une fois apparu, un événement reste affiché dans l'emprise de la carte. Une forme de mémoire des événements passés est donc maintenue visuellement, ainsi qu'une indication de la chronologie *localement* : en effet, la chronologie déductible de l'empilement n'est valable qu'à l'intérieur de la pile. Le fait de conserver à l'écran tous les événements permet cependant d'afficher des lignes de causalité même lorsqu'un événement a des conséquences à des dates très éloignées de sa

¹²⁰ Pour une meilleure compréhension de cette animation, nous invitons le lecteur à visualiser la vidéo téléchargeable à l'adresse http://lig-membres.imag.fr/saintmar/docs/demo_carto_evts_inondations.zip

propre apparition. Cela permet, en outre, de bénéficier d'une vue globale sur l'ensemble de la situation à la fin de l'animation (comme illustré dans la vue donnée par la Figure 36).

Pour aider à la perception des changements, l'animation intègre différents indices visuels qui orientent l'attention du lecteur sur les éléments à observer. Par exemple :

- le clignotement de la date (affichée en bas à droite de la carte) lorsque l'on passe à un autre jour et des symboles des événements au moment où ils apparaissent facilite leur détection, grâce à la saillance induite par l'effet de flash (Khalil et al., 2005; Shipley et al., 2013).
- Le changement de date est également mis en évidence dans la ligne de temps par le passage à un fond bleu de la portion dédiée à ce jour (voir Figure 37). Le déroulé du temps (non proportionnel à la réalité des événements) est signalé par l'avancée d'un petit carré vert sur l'axe horizontal, fur et à mesure de l'animation (dans la Figure 37, il est par exemple placé au point correspondant au quart de la journée du 31/12/1947).
- Le déplacement des événements, de leur localisation initiale vers le cadre blanc d'une pile chronologique, rend explicite le fait que nous optons volontairement pour leur délocalisation (*i.e.* le déport dans la carte) à des fins de lisibilité globale de la carte (voir section 4.3.6).
- L'ajout de figurés jaunes parcourant les lignes en montre leur orientation, de l'événement cause vers son effet, mais permet également de mettre l'accent sur la dispersion spatiale potentielle des conséquences d'un événement.

Dans cette proposition de géovisualisation, percevoir la réalité des réactions en chaîne n'est toutefois possible que par la visualisation complète de l'animation associée à une situation et/ou en cliquant successivement sur les événements selon une démarche qui consiste à « suivre » les relations causes/conséquences. L'expérimentation menée auprès de 22 experts du domaine ferroviaire (voir (Saint-Marc, 2017), Chapitre VII) confirme la difficulté qu'il y a à lire les relations de causalité et à reconstruire leur enchaînement sous la forme d'effets dominos. Cela nous conduit à repenser l'approche de restitution de ces informations.

4.4.3 Améliorer la représentation des effets dominos

La méthode de représentation des relations de causalité doit donc être améliorée, de façon à permettre une lecture plus facile et plus efficace des processus complexes que sont les effets dominos. Parce que ces effets ne peuvent être perçus que globalement, une vue d'ensemble sur les relations de causalité manque. Une proposition en ce sens consiste à afficher, lors de la sélection d'un événement, l'ensemble de la chaîne de causes l'ayant provoqué (directement ou indirectement) et/ou de l'ensemble de sa chaîne de conséquences¹²¹. La Figure 39 illustre ce procédé : l'événement étudié est le deuxième, en partant du haut, dans la pile chronologique la plus à gauche. Il concerne les dommages sur un ouvrage d'art, dont on voit, en suivant les lignes rouges, qu'ils résultent de l'enchaînement de quatre phénomènes naturels (deux sont placés dans l'encart cartographique en haut à droite de l'image), puis d'un premier incident portant probablement sur ce même ouvrage (existence du même symbole d'ouvrage d'art en haut de la pile chronologique, correspondant à la même localisation). Une analyse similaire peut être menée pour explorer la chaîne des conséquences des dommages sur cet ouvrage en suivant les lignes noires.

En complément, une nouvelle vue dans l'interface (Figure 40), contenant un graphique en réseau représentant la chaîne de causalités entre les *descriptions* d'événements est proposée suite à une suggestion de participants à l'expérimentation.

¹²¹ Par rapport à la démarche décrite dans la section 4.3.7 où nous construisons des listes de causes et de conséquences, ici l'obtention des informations nécessaires repose sur l'extraction du sous-graphe des relations de causalité (prédicat *idisfer:causes*) dont l'événement observé est le centre, *i.e.* le sous-graphe est celui incluant uniquement les chemins qui mènent à cet événement ou qui en partent.

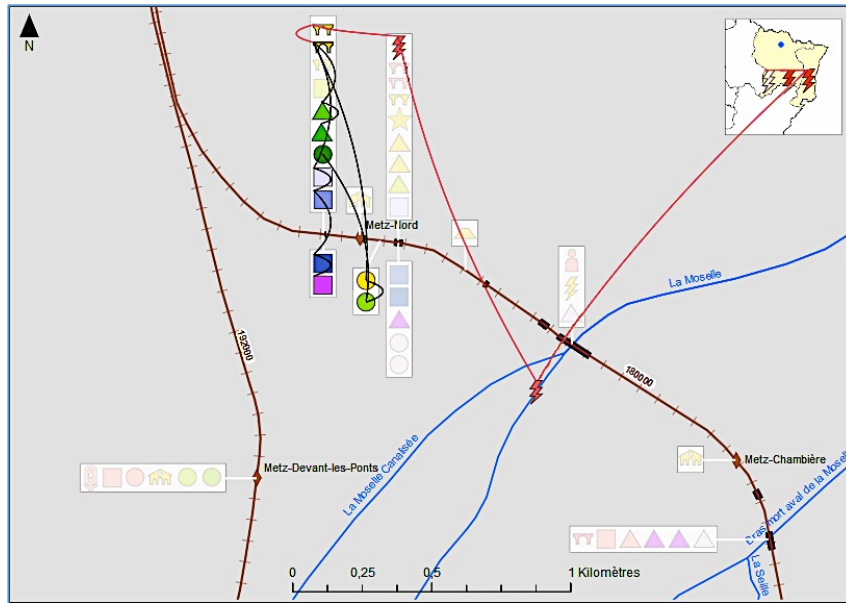


Figure 39 - Exemple d'affichage de l'ensemble de la chaîne de causalité menant à un événement et en partant

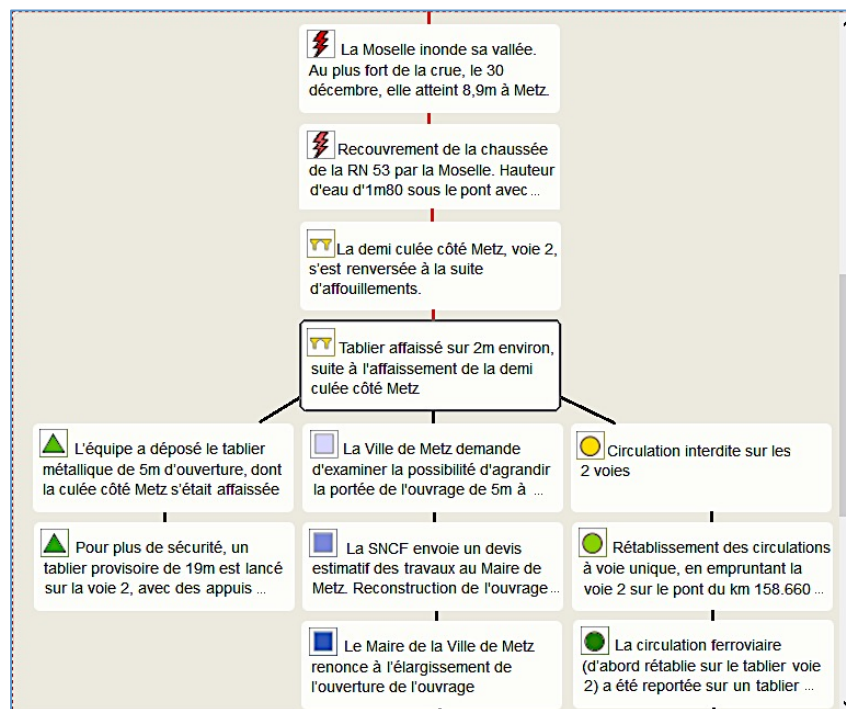


Figure 40 - Exemple d'un graphique en réseau pour chaîne de causalité menant à un événement et en partant

4.5 Synthèse et prolongements

Dans ce chapitre, nous avons exposé certains des travaux réalisés dans le cadre de la thèse de C. Saint-Marc [TH-4]. Pour ce qui concerne le *domaine de la géovisualisation* initialement au cœur de ces recherches, trois propositions innovantes ont été formulées :

- Premièrement, le concept de *piles chronologiques* d'événements, animées au cours du temps, permet de visualiser les événements superposés et de conserver une trace des scènes antérieures de l'animation, prolongeant ainsi la mémoire de travail de l'utilisateur dans sa démarche exploratoire notamment convoquée dans le cas d'un REX¹²². L'interactivité associée à la carte qui inclut ces

¹²² Dans sa thèse, C. Saint-Marc aborde en filigrane plusieurs pans de la cognition, à travers différents mécanismes, déclenchés dans des traitements « simples » de l'information, tels que la perception, ou mobilisés pour des traitements plus

piles est essentielle pour que l'utilisateur puisse réinterroger les événements et les filtrer selon ses objectifs.

- Deuxièmement, la carte donne à voir explicitement les relations de causalité entre événements, sous la forme de lignes qui lient les événements entre eux. En version non animée (visualisation à la demande, au clic sur un événement) ou animée (déroulé de toute la situation), la proposition permet aux utilisateurs d'explorer et d'analyser ces relations et leurs impacts, dans le processus d'endommagement des infrastructures notamment. Une visualisation de la chaîne des relations de causalité permet d'en percevoir la nature de type « effet domino ».
- Troisièmement, la ligne de temps, déclinée en trois versions (à pas de temps régulier, à pas de temps proportionnel au nombre d'événements et synthétique), fournit une vue globale de la répartition et de la dynamique des événements sur l'ensemble de la période étudiée. Les trois formes sont complémentaires et permettent d'accéder à des temporalités différentes pour étudier la situation.

Concernant le domaine de la *représentation des connaissances*, une contribution de la thèse est l'ontologie IDISFER qui revêt à la fois une dimension de *domaine* par la représentation d'un système ferroviaire, une dimension *métier* par la caractérisation des incidents et mesures de réactions au cœur de l'action des services de la SNCF, et une dimension *applicative* dans la possibilité qu'elle donne de gérer l'information historique relative aux situations de crise ayant touché le territoire. Du fait de ces dimensions, l'ontologie est un outil qui, couplé aux propositions de géovisualisation précédentes, contribue potentiellement à augmenter la qualité des expertises de type REX, en donnant à voir différemment l'information. Les premiers tests menés dans le cadre de la thèse visaient à évaluer l'utilité et l'utilisabilité de nos propositions de restitutions géovisuelles des connaissances. Leur impact sur le raisonnement (humain) et le processus de prise de décision des experts reste selon nous à explorer. Je reviens sur ce point à l'occasion de la présentation des perspectives pour mes travaux de recherche (voir section 5.2).

Cette ontologie pourrait également servir à une autre forme de raisonnement (automatique) visant la simulation de situations de crise. Ainsi, en partant d'un phénomène naturel hypothétique, le parcours des relations de type effets dominos identifiées permet de connaître la chaîne des incidents auxquels s'attendre, les perturbations envisageables et les mesures de réactions qu'il faudrait alors prévoir, et ce, pour les différents objets du système ferroviaire. Néanmoins, aller vers un outil de simulation performant requerrait, bien sûr, de coupler la connaissance portée par IDISFER avec, notamment, d'autres sources de données et des raisonneurs spatio-temporels capables de tenir compte de la réalité du terrain et de la dynamique des phénomènes simulés. En prolongeant la réflexion sur ce sujet, je relève une certaine insuffisance dans l'expression des relations de causalité. En effet, avec les prédicats *idisfer:involves* et *idisfer:mayCause* (qui traduisent les relations de causalités théoriquement envisageables) et le prédicat *idisfer:causes* (qui dénote l'observation d'une relation de causalité lors d'une situation avérée), nous simplifions nettement la réalité en occultant les questions relatives au degré de certitude à accorder aux relations, ou encore relatives aux causes multiples intervenant dans des proportions différentes dans le déclenchement d'un événement. Il nous paraît alors pertinent de faire évoluer IDISFER sur ce point en évaluant l'intérêt d'approches exploitant des ontologies probabilistes, floues ou encore possibilistes (voir par exemple (Caglioni and Fusco, 2014) pour une revue de l'intérêt de ces approches pour la connaissance en géographie). Ainsi, les simulations de situations de crise pourraient être élaborées plus finement, sur la base de connaissances moins tranchées mais, au contraire, plus en adéquation avec la complexité des phénomènes spatio-temporels. Dans un but d'opérationnalisation, nous pourrions nous inspirer¹²³, par exemple, des travaux tels que ceux décrits dans (Fusco, 2012). Ils reposent sur une approche probabiliste bayésienne qui permet de produire des modèles pour la création de scénarii de géo-prospection territoriale intégrant l'incertitude. Ce travail adopte une vision systémique du fonctionnement de l'aire d'étude au cours du temps et

élaborés de construction du sens. En cela, ces travaux émergent à l'Action Prospective *Géovisualisation et Cognition* du GDR CNRS MAGIS, animé par S. Christophe et moi-même.

¹²³ Ces travaux ne reposant pas sur une ontologie pour la représentation des connaissances, une adaptation en ce sens est à prévoir.

prend en compte l'expression d'objectifs pour transformer le réseau bayésien en un graphe de décision.

Dans le cadre du travail de C. Saint-Marc, nous avons aussi formalisé des propositions sémiologiques sous forme de règles visant à constituer un premier cadre pour faciliter la réutilisation de nos propositions (pour, de manière immédiate, appliquer la même approche à la visualisation d'autres situations d'inondations répertoriées dans les archives de la SNCF, et, à plus long terme, l'adopter pour d'autres contextes). Ces règles constituent également un premier pas vers l'automatisation de la cartographie des événements spatio-temporels, à partir des récits instanciés dans l'ontologie IDISFER. Dans ce manuscrit, nous prolongeons ce travail en exploitant une approche complètement basée sur des ontologies pour exprimer des équivalents à ces règles. Nous proposons de construire, par une ontologie de *matching* dédiée, une représentation intermédiaire entre la représentation ontologique de connaissances thématiques (et par ce terme, nous couvrons les aspects domaine, métier et applicatif) et la représentation ontologique de ce qu'est une *géovisualisation*. Elaborée à partir des concepts de l'ontologie thématique *IDISFER* et de l'ontologie *GEOVIS*, l'ontologie *GVR Matching* vise à faciliter la mise à disposition des informations sous une forme adaptée en vue de l'implémentation.

L'ontologie *GEOVIS*, à peine ébauchée dans ce document, vise à organiser, à un haut niveau d'abstraction et de manière consensuelle, les concepts qui participent de la constitution d'un outil de géovisualisation, vu comme « *un outil dynamique et interactif pour afficher des données géographiques, des modèles de données, des cartes, des images, et des données externes géo-localisables, permettant à un utilisateur de raisonner spatialement et/ou temporellement sur ces représentations, et d'inférer des connaissances sur le territoire ou le phénomène représenté, en offrant la capacité d'interagir avec une ou plusieurs dimensions du phénomène.* » (extrait de (Christophe, 2017), p5). Les principes d'élaboration de l'ontologie *GVR Matching*, présentés dans ce chapitre, ont été déterminés de manière empirique sur le cas de l'ontologie *IDISFER*. Ces propositions doivent donc être considérées comme un travail initial, presque exploratoire, que nous poursuivrons dans le cadre de nos futures recherches.

« La certitude et le mystère sont pour le sentiment ;
la clarté et l'incertitude pour le raisonnement. »

Antoine de Rivarol (1753 - 1801)

C **onclusion** 5

A travers ce document, j'ai présenté, et revisité, certaines des activités de recherche que j'ai menées depuis la fin de ma thèse. Ce chapitre de conclusion est, comme il se doit, l'occasion de dresser un bilan. J'analyse les apports des travaux du point de vue de la représentation des connaissances. Je me livre ensuite, à l'incontournable, mais tout autant enthousiasmant, exercice d'ouverture sur de futures recherches.

5.1 *Bilan*

Le champ de la représentation *géovisuelle* de connaissances a essentiellement été abordé, jusqu'ici dans ce document, à travers le travail lié à la thèse de C. Saint-Marc (voir Chapitre 4). La section 4.5 (voir page 87) donne déjà des éléments de synthèse, aussi allégeons-nous ce bilan en ne reprenant pas cette thématique. Il en sera toutefois à nouveau question dans la section 5.2, consacrée à nos perspectives de recherche.

Nous orientons donc nos propos sur la représentation (conceptuelle) des connaissances. Dès la première thèse que j'ai co-encadrée, celle de M. Bilasco [TH-1], les approches sémantiques pour la représentation des connaissances ont dessiné le contexte de mes recherches, mais c'est avec les thèses de A.D. Miron [TH-2] et de A. Hombiat [TH-3] que j'ai rejoint des questionnements relevant de la communauté géomatique (voir préambule, page 1). Dans ce manuscrit, mon activité de recherches a été décrite en mettant principalement l'accent sur les trois thèses, dont j'ai également assuré le co-encadrement, et qui ont suivi celles évoqués ci-dessus. Le point commun, qui a motivé le choix de présenter ces travaux en particulier, est celui de la représentation de connaissances relatives à l'évolution, une notion que l'on retrouve dans :

- les différentes versions de découpages territoriaux (Thèse de C. Bernard [TH-6], voir Chapitre 2),
- les trajectoires de vies des personnes (Thèse de D. Noël [TH-5], voir Chapitre 3) et
- les récits de situations de crise (Thèse de C. Saint-Marc [TH-4], voir Chapitre 4).

Ces travaux ont permis d'apporter des réponses à des problèmes de représentation de connaissances dans des contextes applicatifs particuliers et différents, mais pour lesquels il s'agit toujours de favoriser la compréhension de phénomènes passés et/ou la prise de décision. Il s'agit

clairement de travaux de recherche appliquée, dans lesquelles les ontologies développées sont des solutions à des problèmes concrets. Les approches ontologiques proposées répondent aux objectifs que nous avons évoqués en (voir Introduction, page 7) : garantir une certaine généralité dans ces ontologies spécifiques, intégrer dimensions descriptive et explicative, supporter l'imperfection. Parce que les objets d'études et les usages diffèrent, les réponses apportées sur ces trois points diffèrent également dans les approches.

Généricité - Les ontologies TSN et TSN-Change sont des ontologies génériques pour la gestion des nomenclatures territoriales, applicables à différents découpages de territoires portés par des institutions gouvernementales ou gestionnaires de données. Ces ontologies sont le résultat d'une démarche *data-driven* (au sens des approches symboliques, voir la discussion page 7), issue de l'étude des données de nomenclatures existantes, dont la NUTS en particulier ; pour cette dernière, nous avons décrit et mis à disposition dans le Web des LOD, les instances conformes aux ontologies TSN et TSN-Change, de toutes les versions existantes. Des travaux en cours sur le découpage du territoire australien montrent l'applicabilité des propositions à d'autres contextes et témoignent de leur généralité.

Dans l'approche LTOP, la généralité est dans la démarche à base de patrons permettant de construire ses propres ontologies applicatives. L'approche a été initiée par le besoin de collecter des données mettant en exergue les migrations résidentielles, pour un contexte d'études rapprochant sociologie et urbanisme. Elle offre cependant la possibilité de façonner aisément ce qu'est une trajectoire de vie pour les besoins d'une étude en bénéficiant de la capacité à décrire des dimensions multiples et d'inclure des facteurs explicatifs. La généralité de ce travail apparaît également dans la possibilité offerte de représenter, sur la même base conceptuelle, non seulement les faits avérés mais aussi les aspirations (passées ou pour l'avenir) des individus.

Enfin, l'ontologie IDISFER est en elle-même une ontologie générique qui offre une description très complète des concepts du domaine ferroviaire. Elle est aussi une ontologie métier en cela qu'elle décrit les phénomènes pouvant potentiellement impacter ce système, servant à l'instanciation d'ontologie d'application pour retranscrire la connaissance détenue dans les données d'archives (comme réalisé pour cinq inondations ayant touché le réseaux français). Enfin, la typologie des imperfections proposées dans cette thèse, versée dans le Web des LOD, est une ontologie générique qui peut être exploitée dès qu'il s'agit de qualifier les connaissances sur ce point.

Dimensions descriptive et explicative - Dans ces travaux, nous avons contribué à représenter les connaissances avec une visée *descriptive* des faits en rapport avec l'évolution observée et une visée *explicative* de ceux-ci. Le Tableau 1 donne, pour chacune des trois approches décrites, les éléments de représentation des connaissances associés à ces points de vue descriptif et explicatif. Les approches partagent en effet, le fait d'intégrer chacune à leur manière ces deux dimensions dans une représentation, ainsi plus riche, de l'évolution.

Tableau 1 – Visées descriptive et explicative des connaissances représentées dans les trois approches présentées

	Visée Descriptive des Faits	Visée Explicative des Faits
Découpages territoriaux évolutifs (Chapitre 2)	<ul style="list-style-type: none"> ▪ Structuration hiérarchique des territoires ▪ Succession de versions d'un découpage 	<ul style="list-style-type: none"> ▪ Caractérisation des changements ▪ Liens inter-niveaux des changements ▪ Référencement de causes économiques, politiques, sociétales, etc. dans le Web des LOD
Trajectoires de vie personnelles (Chapitre 3)	<ul style="list-style-type: none"> ▪ Multi-dimensionnalité des trajectoires ▪ Succession d'épisodes et événements 	<ul style="list-style-type: none"> ▪ Facteurs explicatifs ▪ Référencement de causes économiques, politiques, sociétales, etc. dans le Web des LOD
Récits de situations de crise (Chapitre 4)	<ul style="list-style-type: none"> ▪ Structure du réseau ▪ phénomènes, impacts, mesures de réaction ▪ Succession d'événements 	<ul style="list-style-type: none"> ▪ Liens de causalité / effets dominos ▪ Référencement de sources (historiques, journalistiques, etc.) dans le Web des LOD

Derrière ces réalisations concrètes, portées par des concepts dédiés dans chaque domaine, nous défendons l'idée qu'il fait sens d'avoir à l'esprit ces visées descriptive et explicative dans une approche de construction d'ontologies dédiés à étayer la compréhension de phénomènes et à la prise de décision. La dichotomie entre ces visées n'est qu'apparente dans nos propos : elle n'a d'autre vocation que de mettre l'emphase sur l'intérêt (méthodologique, en matière d'ingénierie des connaissances) de considérer un domaine selon ces deux points de vue fortement complémentaires : les faits et les éléments pouvant les expliquer. Une des opportunités offertes par le Web des LOD est d'ouvrir sur un monde de connaissances auquel il est devenu très courant de faire référence pour la partie descriptive de la représentation de connaissances (par le recours à des standards de description ou le *linkage* avec d'autres ontologies). Dans nos travaux, nous montrons qu'il s'agit d'une source également pertinente pour renforcer la représentation de connaissances explicatives des phénomènes.

Imperfection – Le travail de C. Saint-Marc (Chapitre 4) est celui dans lequel l'imperfection des données est gérée le plus explicitement par la proposition d'une typologie détaillée des imperfections non présentées ici mais que le lecteur pourra trouver dans (Saint-Marc, 2017). La catégorisation élaborée permet de qualifier précisément la nature de l'imperfection, et, par suite, d'informer l'utilisateur, favorisant ainsi une prise de conscience de la qualité des connaissances manipulées. Cela va dans le sens d'une prise de décision éclairée.

Dans l'approche TSN / TSN-Change, c'est la possibilité même de décrire les changements territoriaux, en ayant qui plus est recours à une typologie fine des changements, qui se veut être une contre-mesure à l'imprécision. La stratégie est donc plutôt de chercher à éliminer le doute par un gain en précision dans les descriptions. Dans ces travaux, une forme d'incertitude peut par ailleurs exister lorsqu'il s'agit de définir ce qui fait l'identité d'une unité spatiale (son code interne à la nomenclature, son nom, sa géométrie, etc.). En l'absence de certitude sur ce point, l'algorithme de détection des changements que nous proposons est suffisamment flexible pour définir de qui détermine cette identité, à l'aide de pondérations associées aux différents composants de celle-ci.

Enfin, l'approche LTOP va de pair avec une démarche de collecte de l'information la plus précise possible. C'est en incitant le contributeur à fournir un maximum d'informations détaillées que nous cherchons à éliminer les imperfections (imprécision et incomplétudes notamment). À défaut d'obtenir d'emblée la plus fine granularité d'information, les organisations hiérarchiques des concepts (spatiaux, temporels, ou thématiques) sont exploitées pour tenter de collecter l'information au niveau de granularité immédiatement supérieur (comme, par exemple, le quartier si la rue n'est pas fournie, le mois et l'année à défaut du jour précis, la CSP à défaut du poste occupé).

Les travaux que nous avons présentés dans ce manuscrit offrent ainsi des prolongements possibles sur différents points spécifiques aux sujets qu'ils abordent et j'entends leur consacrer une partie de mes travaux futurs. Différentes actions de recherche sur lesquelles peuvent ouvrir ces travaux ont été mentionnées dans les sections 'Synthèse et Prolongements' qui concluent chacun des chapitres (voir sections 2.6, 3.5 et 4.5). Dans la section suivante, je présente des perspectives de recherche qui mettent l'accent sur le lien entre représentation de connaissances, géovisualisation et raisonnement humain.

5.2 Perspectives de Recherches

La problématique autour de laquelle je souhaite organiser mes recherches est la suivante : en quoi les techniques de représentation de connaissances et de géovisualisation peuvent-elles contribuer à assister le raisonnement humain, nécessitant l'analyse d'informations imparfaites, dans un contexte de prise de décision ? Comme évoqué dans (Fabrikant, 2001), face à la multitude des formes de restitutions possibles pour l'information géographique, un des défis en géovisualisation est de formaliser la connaissance et les stratégies qui sous-tendent le raisonnement pour mieux répondre aux besoins de l'utilisateur en situation de prise de décision.

5.2.1 Cadrage et Enjeux

Deux cadres applicatifs peuvent être introduits pour illustrer les enjeux de ces recherches. Dans les travaux de C. Saint-Marc (voir Chapitre 4), nous avons mis l'accent sur la représentation des récits relatifs à des inondations et expliqué que ce travail ambitionne de revisiter la façon dont sont conduits

les REX : l'expert pourra désormais formuler des décisions (de natures diverses : recommandations de travaux, préconisations pour réorganiser des services, par exemple) en se basant sur de nouvelles formes de représentations (pile chronologiques, *timelines*, etc.) pour analyser des événements survenus tout au long de la situation de crise. Or, nous n'avons pour l'instant que peu investi la façon dont les informations présentées (et plus précisément représentées *visuellement*) sont comprises et intégrées dans le raisonnement de l'expert dans un cadre de REX.

Le second exemple est celui du secours en montagne auquel nous nous intéressons dans le cadre du projet ANR CHOUCAS¹²⁴ (Olteanu-Raimond et al., 2017) dont l'objectif est de proposer des modèles, méthodes et outils facilitant le raisonnement du secouriste (ici membre du PGHM¹²⁵) qui reçoit un appel de détresse et doit localiser la victime pour envoyer les secours. Il se base pour cela sur l'analyse d'un ensemble d'indices fournis par le requérant¹²⁶, décrivant notamment des objets géographiques (éléments du paysage, points de repères, itinéraires empruntés, ...). Globalement, un processus de raisonnement, notamment spatial, est opéré par le secouriste pour tenter de déduire la localisation la plus probable de la victime : il formule des hypothèses à partir des premières informations fournies par le requérant, et au cours du dialogue, il les affine progressivement, ou les remet en question, à la lumière de son analyse des indices, de la confiance qu'il leur accorde, et de sa connaissance experte. Plusieurs éléments compliquent cette tâche de localisation. Du côté de la victime, une faible connaissance de la montagne, une représentation approximative de la géographie des lieux, des imprécisions dans la description ou le nommage de l'itinéraire emprunté, mais aussi la peur, l'angoisse ou encore la douleur, font que la description de sa localisation est souvent imprécise et pas toujours fiable. Du côté du secouriste, l'hétérogénéité des données sur lesquelles il s'appuie peut constituer un frein dans le processus, du fait de nombreuses données multi-sources (données institutionnelles ou communautaires) et multi-formats (cartes, topo-guides de randonnées en format papier, données vectorielles), qui offrent des informations parfois non immédiatement identifiables comme étant les mêmes.

Ces deux exemples présentent des dénominateurs communs. D'une part, les informations manipulées dans l'analyse et le raisonnement opérés sont de différentes natures (dont spatiales et/ou temporelles) et caractérisées par une imperfection qui prend différentes formes (incertitude, l'imprécision, incomplétude, etc.) (Saint-Marc, 2017). D'autre part, un outil de géovisualisation est proposé comme support de l'analyse et du raisonnement de l'expert. Il est donc particulièrement important que la gestion de ce caractère imparfait ne soit pas négligée dans la géovisualisation qui l'intègre (Griethe and Schumann, 2005). Le cas de la localisation en montagne présente des difficultés supplémentaires par rapport au cas de type REX, en cela que la situation à analyser est en cours et que le raisonnement doit mener à une décision rapide (*où* envoyer les secours ?). Notons néanmoins que les enjeux économiques et sociétaux pour ces deux situations ne sont pas neutres, puisque des moyens matériels et humains importants peuvent être engagés, avec des coûts et des risques parfois non négligeables.

Pour illustrer les pistes de recherche sur lesquelles je souhaite m'engager, je reviens dans ce qui suit au projet CHOUCAS. Les questions soulevées trouvent parfois cependant un écho au delà de ce cas d'application. Plusieurs problématiques parmi celles introduites font en partie l'objet de la thèse de M. Viry [TH-8].

5.2.2 Des stratégies de raisonnement à formaliser

L'observation des pratiques que nous avons menée au sein du PGHM (Villanova-Oliver, 2018) montre qu'il n'existe pas vraiment de procédure de raisonnement standardisée ni formalisée. Des approches heuristiques, basées sur l'expérience – dans la gestion des alertes, connaissance du terrain concerné, connaissances tacites, *etc.* –, mais aussi l'intuition, sont employées par l'opérateur en charge de localiser la victime. La méthode globale (appelée stratégie de décision, voir par exemple (Payne et

¹²⁴ CHOUCAS (ANR-16-CE23-0018), <http://choucas.ign.fr/> - *Choucas* est l'indicatif radio des hélicoptères de la gendarmerie, utilisés, entre autres, pour le secours en montagne. Voir Annexe 2 page 109 pour une présentation du projet.

¹²⁵ Pelotons de Gendarmerie de Haute-Montagne

¹²⁶ Terme consacré pour désigner la personne déclenchant les secours, qui peut ou non être la victime, et peut ou non être à proximité de celle-ci

al., 1993)) selon laquelle des décisions sont prises par l'être humain, repose sur une séquence d'opérations de recherche dans un espace de solutions. Il est reconnu que les caractéristiques de la situation dans laquelle se fait la prise de décision influent sur la stratégie de décision adoptée par un agent : elle peut se révéler plus analytique qu'heuristique par exemple (Canellas et al., 2015). Ainsi, dans notre contexte, les caractéristiques de la situation d'alerte (liées, entre autres, à l'urgence de la situation, à la crédibilité des sources d'information disponibles, à la qualité des indices, au caractère séquentiel de leur acquisition, etc.) ont certainement un impact sur le raisonnement opéré, mais elles sont à ce jour non formalisés, non verbalisés, implicites.

Formaliser les étapes du raisonnement et les données nécessaires à celles-ci constitue un objectif scientifique dont des bénéfices pourront être retirés à, au moins, trois niveaux : *i*) analyser plus facilement les raisonnements opérés à des fins d'amélioration des pratiques si possible, *ii*) faciliter la transmission des savoirs lors de la formation de nouveaux secouristes par exemple, *iii*) opérationnaliser une partie du processus de raisonnement dans un outil dédié à l'activité de localisation. Il s'agit donc de capitaliser l'expertise et l'expérience des secouristes, à l'image par exemple des travaux de (Abi-Zeid et al., 2010) dans le cadre de la recherche d'avions disparus. Dans ce cadre, des approches, généralement exploitées dans le domaine de l'entreprise seront revisitées (par exemple (Ermine, 2003; Saad et al., 2009)), pour produire des modèles de données et de connaissances adaptés à notre contexte. Nous serons confrontés à la complexité du processus de raisonnement (du fait des caractéristiques évoquées plus haut), mais aussi parce qu'il n'existe pas *un* processus de raisonnement mais autant de processus qu'il existe de secouristes, de requérants, de situations justifiant l'appel d'urgence. Produire une formalisation ayant le 'bon' niveau de description pour atteindre les objectifs d'analyse des pratiques, de transmission des savoirs et d'opérationnalisation est un verrou scientifique auquel nous tenterons de répondre. Une analyse de l'état de l'art sera menée pour identifier les modèles existant à même de répondre à tout ou partie de nos besoins. Seront, entre autres, étudiés : les modèles de comportement issus d'approches cognitives, de type *Belief-Desire-Intention* (BDI, (Bratman, 1999)) ou encore *Recognition-Primed Decision* (RPD, (Klein, 1998)) et les modèles de tâches, souvent en lien avec des approches ergonomiques (voir, à titre d'exemple, (Burkhardt et al., 2003; Couix et al., 2005; Le Pors, 2010) pour une revue des modèles).

Deux axes d'études nous intéressent plus particulièrement et, de par leur lien avec la géovisualisation et la gestion de l'imperfection, font l'originalité de cette partie du projet :

- la mise en évidence et la formalisation des différentes opérations d'exploration et d'exploitation des *composants d'une géovisualisation*, en particulier des données cartographiques, dans la stratégie de décision mise en œuvre. Ici, une démarche d'observation et de mesure des comportements, essentiellement basée sur des techniques d'*eye-tracking* (Holmqvist and Andersson, 2017; Poole and Ball, 2005) nous permettront d'objectiver le rôle des outils supports de l'analyse géovisuelle dans la stratégie de décision.
- l'impact du caractère imparfait des indices collectés sur la stratégie de raisonnement, sachant qu'il convient de tenir compte du jugement que porte le secouriste sur un indice (imparfait ou non). Au delà de travaux existants sur ce thème (voir par exemple (Canellas et al., 2015)), nous nous intéresserons au fait que l'encodage du caractère imparfait, est réalisé par le secouriste lui-même à partir de ce que lui livre le requérant, et de comment il le livre : avec aplomb, en hésitant, par exemple sur des évaluations de durée ou de distance, en se contredisant, etc. La qualification de l'imperfection de la donnée est faite sur le moment et elle est subjective : cela amène-t-il à des stratégies différentes, plus ou moins prudentes, et en quoi cela influence-t-il l'efficacité (en termes d'exactitude et de rapidité) avec laquelle est trouvée la localisation finale ?

5.2.3 Un raisonnement étayé et traçable

La suite logique de la représentation du raisonnement est *d'essayer de le faciliter voire de l'améliorer*. Nous ne visons pas à ce jour l'automatisation complète d'un processus de raisonnement venant se substituer à celui secouriste. Il ne s'agit donc pas de définir un système de décision, qui, par exemple à partir de règles, appliquées à un ensemble de faits, produirait automatiquement la solution, c'est-à-dire identifierait la localisation recherchée. Il est *i*) à ce stade, prématuré de l'envisager, compte-tenu de la nécessité de mener en amont les recherches évoquées en section 5.2.2 ; *ii*) même à plus long terme, cela resterait certainement difficile à réaliser du fait des caractéristiques de l'espace

de problème (imperfection des données, singularité de chaque cas, prédominance de l'expertise humaine et de l'intuition dans le traitement) ; *iii*) dans tous les cas, cette perspective n'est pas souhaitée par les secouristes qui veulent garder la main sur l'outil.

Pour autant, nous relèverons le défi d'une certaine amélioration du processus de raisonnement via des outils dédiés. Trois pistes de travail sont identifiées qui visent à étayer le processus et à en assurer une forme de traçabilité. Les deux premières requièrent une réflexion en termes de conception de composants d'interface, que nous avons amorcée à titre expérimental dans le cadre du Master de Matthew Sreeves (voir page 116, (Sreeves, 2018)).

Dans un premier temps, un outil facilitant le *recueil et la qualification des indices en termes d'imperfection et de crédibilité* est à envisager. Il convient de permettre une acquisition flexible de ces éléments, signifiant la possibilité d'ajouter facilement au fil de l'échange toute information, associée si besoin à une forme d'imperfection (incertitude, imprécision, etc.), mais aussi pondérée par le crédit qu'accorde le secouriste à cette information. Il s'agit par exemple ici de représenter l'information suivante : « je suis à environ 400 mètres d'un lac à vol d'oiseau ». L'imprécision est ici sur la distance au lac et le secouriste peut choisir de traduire cela par un intervalle plus ou moins grand, mais aussi, décider d'accorder plus ou moins de crédit à cette donnée globalement. Ces choix sont faits sur la base ce que le secouriste a ressenti dans les échanges précédents avec le requérant et relèvent souvent de l'implicite ; la qualification de l'imperfection et du jugement peuvent évoluer au cours de l'échange. La flexibilité à l'acquisition vise à supporter ces évolutions. Un indice ainsi qualifié (en termes d'imperfection et de crédibilité) doit être exploitable (i.e. mobilisable) à tout moment du processus de prise de décision, signifiant pour le secouriste la capacité d'inclure /exclure un indice s'il le souhaite et de voir l'impact sur les zones de localisation possibles.

D'autre part, dans le prolongement de la possibilité donnée de tenir compte ou non d'un indice, une attention particulière sera portée au fait que les processus de raisonnement nécessitent parfois de maintenir plusieurs hypothèses en parallèle. Pouvoir aisément, par le biais d'une interface, formuler des hypothèses concurrentes, les étayer et les affiner, les remettre en cause, etc. est une valeur ajoutée pour le secouriste. Cela est d'autant plus intéressant si, de plus, les résultats associés aux hypothèses sont visuellement comparables.

La Figure 41 illustre un premier prototype permettant de visualiser les zones de localisations possibles (en rouge dans la carte) étant donnés trois indices collectés à ce stade du dialogue entre le requérant et le secouriste. Le tableau en haut à gauche permet de voir ces indices :

- une situation estimée à 2km de la Croix de Chamrousse qui se traduit par le grand cercle orange dans la figure (le petit cercle au centre donne la position du lieu-dit la Croix de Chamrousse)
- une position évaluée à 100 à 200m autour d'un lac, traduite par des *buffers* encerclant les lacs de la zone
- une situation à 100m d'une remontée mécanique, là aussi traduite par des buffers englobant les lignes symbolisant ces objets.

Les zones de localisation probables sont ici déterminées simplement par intersection des *buffers*.¹²⁷

Ce prototype ne gère pas tous les aspects permettant de caractériser un tel problème de localisation à base d'indices. Notamment, la question du jugement de crédibilité n'est pas supportée autrement que de manière binaire (l'indice est retenu ou non par le secouriste, voir Figure 42) ; de même l'éventail des relations de positionnement par rapport à un objet référence (en référentiel égocentré ou allocentré) ou encore les indications de nature temporelle ne sont pas gérés.

Néanmoins, il permet de facilement modifier les paramètres de distance et d'en voir l'impact sur la carte, ici par des buffers plus ou moins larges autour de l'objet auquel ils se rapportent. De plus, par la simple activation des cases à cocher appropriées dans la partie « scénarios », le secouriste peut suivre jusqu'à trois hypothèses en parallèle, en visualisant aisément quels indices étayaient quels scénarios.

¹²⁷ Dans le cadre du projet CHOUCAS, le calcul des zones de localisation probables sera, à terme, plus complexes et pris en charge par le COGIT pour tenir compte notamment, par exemple de la notion d'intervisibilité ou encore de méthodes avancées pour gérer les zones floues plus finement.

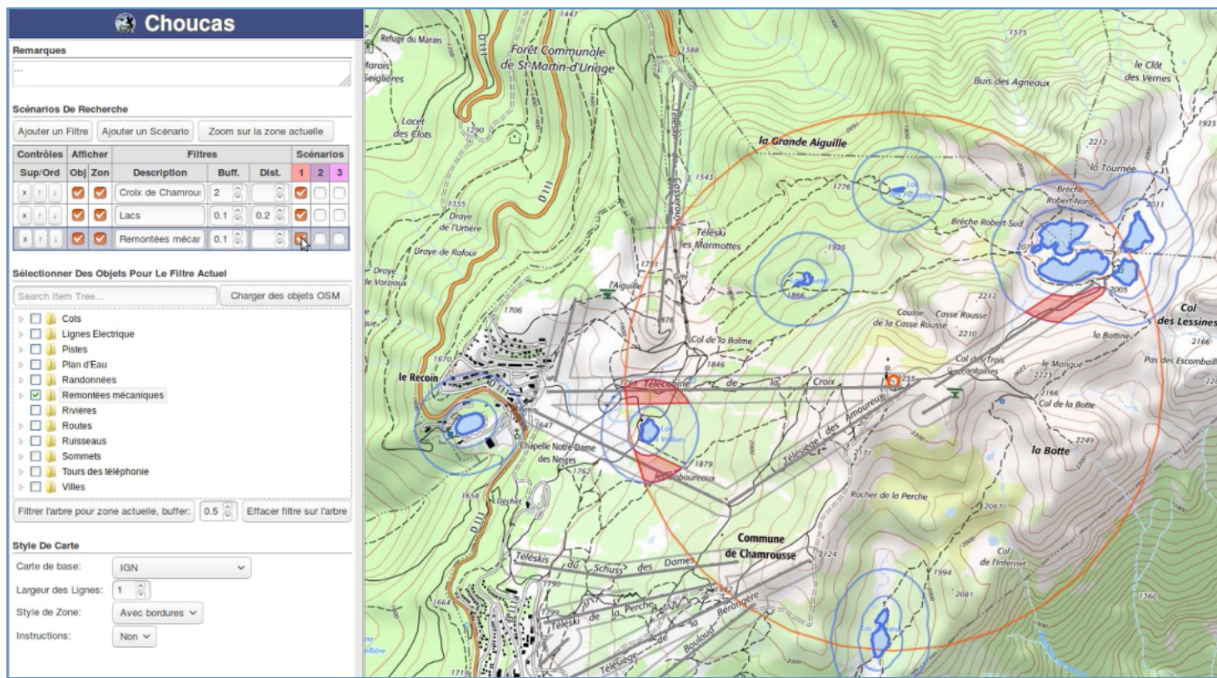


Figure 41 - Prototypage d'une interface d'acquisition d'indices avec imperfection

Exemple de situation caractérisée par 3 indices, menant à la détermination de 3 zones possibles de la localisation

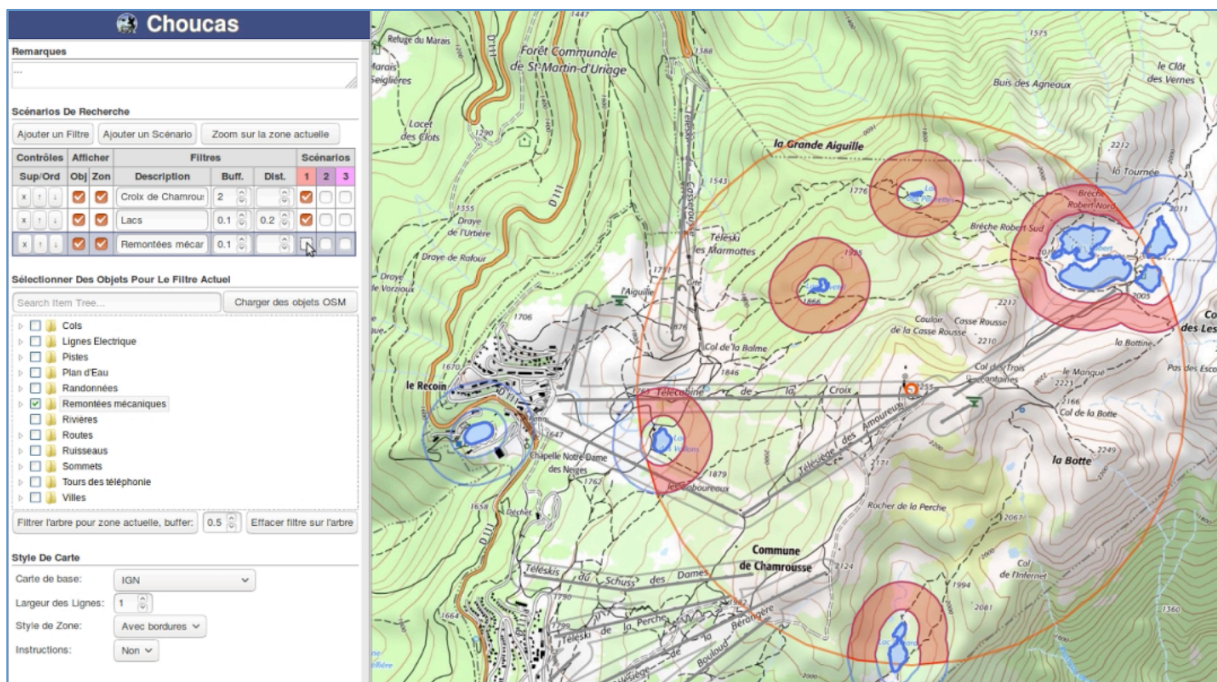


Figure 42 – Illustration de la désactivation d'un indice pour ne plus en tenir compte dans le raisonnement

Désactivé, l'indice relatif aux remontées mécaniques n'est plus considéré dans l'hypothèse. Plus de zones de localisations possibles sont compatibles avec les indices conservés dans le raisonnement.

Enfin, notre troisième axe de recherche pour améliorer le processus de raisonnement passe par une contribution visant le déroulement du dialogue. L'objectif visé est de suggérer au secouriste la prochaine question à poser au requérant pour récolter l'information qui permettra de converger au plus vite vers une localisation suffisamment précise. Cet objectif requiert d'expertiser l'état de la connaissance au temps t et de déterminer quelle(s) information(s) obtenir pour amener cet état de connaissance à un « meilleur » niveau, c'est-à-dire qui permet, par exemple, d'infirmer une hypothèse de localisation ou, idéalement, de la confirmer et d'en réduire l'étendue. Parmi différentes approches

permettant de répondre à cet objectif, celle proposée par (Codet de Boisse, 2013) qui combine raisonnement par analogie et raisonnement déductif est prometteuse, mais demande à être revisitée dans un contexte de localisation en situation d'urgence et caractérisé par des données imprécises ou peu fiables. Les travaux de (Le Pors, 2010) reposant sur de la simulation à base d'agents pourront également être repositionnés dans ce même contexte pour en évaluer l'intérêt. Dans ce travail, je suis particulièrement intéressée par le potentiel qu'il y a à raisonner sur la connaissance, notamment pour rechercher ce qui permettrait de maximiser la différence entre deux états. A titre d'exemple, exploiter la connaissance pour identifier que la nature de la végétation (des sapins versus des feuillus) est un critère discriminant entre deux zones possibles de localisation est un élément permettant d'orienter la prochaine question à poser.

5.2.4 Une sémiologie utile, utilisable et non intrusive

La définition d'une sémiologie permettant la prise en compte du caractère imparfait des données et de l'incertitude dans le raisonnement est un dernier axe de recherche. L'utilité de visualiser le caractère imparfait des données est reconnu dans la littérature depuis longtemps (Beard and Mackaness, 1993; Buttenfield, 1993), mais peu d'environnements de géovisualisation l'intègrent efficacement (Saint-Marc, 2017). De plus, face au déluge des données, notamment en lien avec le Web des LOD, caractérisées par une grande hétérogénéité, donner à voir l'imperfection – ou, plus exactement, toutes les formes d'imperfection – est une question exacerbée (Kinkeldey et al., 2014; MacEachren et al., 2012). Dans les travaux en lien avec cette question, l'effort est par exemple porté sur la proposition de variables visuelles dépendant des différents types d'imperfection (Saint-Marc, 2017) et du type de données spatiales (vectérielles, raster) (Arnaud and Davoine, 2011; Buttenfield, 1993; MacEachren et al., 2012). Pour (Beard and Mackaness, 1993; Kinkeldey et al., 2014), il convient de fournir des modalités d'interaction avec les représentations des imperfections afin de mieux les comprendre. Dans (Kinkeldey et al., 2017), une revue des travaux explorant la façon dont différentes formes de visualisation de l'incertitude influencent la prise de décision est proposée. Nous proposons d'aborder également cette question dont un des enjeux est de ne pas interférer dans le processus d'analyse et de raisonnement, mais bien de trouver la représentation adéquate. Ainsi, la superposition d'emprises graphiques, évoquée en section 4.3.6, reste à solutionner pour la représentations des indices et zones de localisation incertaines qui se chevauchent. En ce qui concerne la transcription visuelle de l'imperfection associée à un élément (indice ou une zone), il s'agira de penser une sémiologie évitant toute sur/sous-représentation de l'imperfection, ceci pouvant être totalement contreproductif puisque conduisant à ignorer l'élément, ou, au contraire, lui accorder trop d'importance dans la prise de décision.

« [...] la clarté et l'incertitude [sont] pour le raisonnement » (Antoine de Rivarol), sous-entendu raisonnement *humain*. Le projet de recherche que j'ai présenté est motivé par le fait qu'il convient, tant dans le champ de la représentation des connaissances que dans celui de la géovisualisation, de continuer à proposer des approches, des méthodes et des outils qui vise à favoriser la première – la clarté du raisonnement – sans ignorer la seconde – l'incertitude dont il porte l'empreinte. A mon sens, des réponses peuvent être trouvées là où se rejoignent ces deux champs de recherche.

Annexe 1 - Projets ESPON Database (phase I) et M4D (Multi-Dimensional Database Design and Development – phase II)

Ces projets (2007-2015), financés par le Programme ESPON¹²⁸ de la Commission Européenne, ont porté sur la conception et le développement d'une infrastructure de données spatiales multidimensionnelles et évolutives.

En matière de production d'informations statistiques, les fournisseurs sont incités à diffuser conjointement les données et leurs métadonnées. En effet, connaître la source des données ou la description de la méthodologie employée pour la construction d'un indicateur permet aux utilisateurs d'appréhender et d'interpréter correctement les phénomènes statistiques décrits, en particulier dans le contexte de comparaisons internationales.

Partant du constat que les métadonnées ont un rôle primordial, les consortiums des projets *Database* puis *Multi-Dimensional Database Design and Development* (M4D) ont œuvré à la construction d'une infrastructure logicielle pour la gestion d'indicateurs statistiques à références spatiale et temporelle¹²⁹. L'enjeu consistait, entre autres, en l'harmonisation et la mise à disposition de jeux de données produits par des fournisseurs multiples contribuant à alimenter le programme ESPON en données sur le territoire européen. Plusieurs axes de contributions peuvent être retenus.

Contribution à l'expression de métadonnées pour les jeux de données statistiques

Ces métadonnées doivent reposer sur un format structuré et partagé, pour autoriser un usage interopérable entre différents systèmes d'information géographique. INSPIRE, une directive européenne pour le partage et la diffusion de l'information géographique, préconise à cet effet l'usage d'un standard pour les données à référence temporelle et spatiale : la norme ISO 19115. Parce qu'elles se présentent le plus souvent comme des ensembles datés de nombres associés à des unités territoriales, les statistiques socio-économiques sont des informations à références spatiales et temporelles qu'il serait légitime de décrire par des métadonnées au standard ISO 19115. Alors que la norme ISO 19115 a été largement étudiée et adoptée par les différents producteurs et usagers de données à références spatiales tels que l'Agence Européenne de l'Environnement et le Joint Research Center, les acteurs du système de collecte de données socio-économiques se tiennent à l'écart de cette norme. Une des raisons est sans doute le manque d'adéquation de la norme ISO 19115 vis à vis des spécificités de l'information statistique. Parmi elles, la structuration particulière, composite et hétérogène, de l'information statistique est à souligner. On raisonne souvent avec un jeu de données statistiques regroupant généralement une collection d'indicateurs différents, mesurés chacun de façon spécifique sur un ensemble d'unités territoriales, à différentes dates. Il existe donc une réelle complexité dans le fait d'associer des métadonnées expressives et exploitables à une telle structure. Un standard et un modèle pour l'échange de données statistiques, appelé SDMX (pour *Statistical Data Model eXchange*) existait déjà lors de nos travaux, insufflé par le domaine bancaire et de celui des assurances. Si SDMX prenait mieux en compte certaines spécificités de l'information statistique que l'ISO 19115, il n'offrait alors cependant pas de fonctionnalités aussi avancées que l'ISO 19115 pour la gestion de l'information spatiale. Ce standard qui mixait, dans une même structure, données et métadonnées, et exigeait des utilisateurs de concevoir et publier le schéma structurant données et métadonnées, était d'une part encore émergent, et, d'autre part, peu opérationnel pour des utilisateurs qui ne souhaitaient pas reformater leurs données dans une structure complexe. Nous avons donc étudié, dans un premier temps, l'adéquation de la norme ISO 19115 pour l'information statistique. Si l'intérêt d'utiliser cette norme pour décrire l'information statistique à composantes spatiale et

¹²⁸ ESPON - European Observation Network for Territorial Development and Cohesion

¹²⁹ voir <https://www.espon.eu/tools-maps/espon-2013-database>

temporelle était confirmé, nos conclusions mettaient aussi en exergue le besoin de l'adapter pour mieux en gérer les spécificités. Nous en avons donc proposé une extension, qui prend la forme d'un profil de la norme, pour la rendre opérationnelle, et ainsi promouvoir son utilisation auprès des acteurs du système d'information pour les données statistiques : producteurs, distributeurs, usagers, etc. Le modèle proposé permet d'appréhender les particularités de l'information statistique territoriale (Plumejeaud et al., 2010). Adopté et respecté par les producteurs de données, le modèle offre le potentiel de contribuer à l'harmonisation, la conformité et la qualité des données recueillies.

Contribution à la qualité des données

Par ce travail, nous avons contribué à une meilleure expression de la qualité des données dite *externe* car elle est avant tout destinée à un utilisateur qui peut, grâce aux métadonnées, mieux juger de l'intérêt de la donnée en regard de ses besoins. Une autre qualité associée aux jeux de données par la norme ISO 19113, dite *interne*, vise à évaluer la qualité intrinsèque du jeu de données comme, entre autres, la confiance qui peut être accordée aux différentes valeurs d'indicateur associées aux unités territoriales. Il s'agit par exemple de déterminer si une valeur semble bien traduire la réalité, si elle paraît plausible dans un contexte donné. Souvent affaire d'experts, des méthodes de détection de valeurs extrêmes (ou aberrantes) peuvent être utilisées à ces fins. Il arrive que les différentes méthodes ne produisent pas des résultats concordants, essentiellement pour des questions d'hypothèses sur les distributions ou de paramétrage. Un environnement permettant de facilement appliquer et comparer plusieurs méthodes pour un jeu de données peut alors s'avérer utile pour des experts devant juger de cette qualité. Dans cette optique, nous avons proposé un outil appelé QualESTIM (Plumejeaud et al., 2011a)(Plumejeaud and Villanova-Oliver, 2012) qui permet à des experts thématiques de choisir et paramétrer les méthodes de détection à appliquer, et facilite la comparaison des résultats par leur présentation via une interface cartographique interactive que nous avons conçue. De plus, l'environnement permet d'accéder facilement aux métadonnées associées au jeu de données étudié (décrite selon le profil ISO 19115 que nous avons proposé). Dès lors, l'expert peut intégrer dans son raisonnement des connaissances supplémentaires (comme par exemple, des connaissances relatives au territoire étudié, au fournisseur des données, ou encore au mode de production des données). Un rapport sur la qualité interne du jeu de données, conforme à la norme ISO 19115, est exportable.

Contribution à la gestion des flux de données statistiques

L'infrastructure logicielle STedi (Bernard et al., 2014)(Bernard et al., 2017b), pour *Spatio Temporal evolutive data infrastructure*, organise la gestion d'un flux de données statistiques : de la collecte des données et métadonnées à leur restitution, et donc du fournisseur à l'utilisateur. Cette infrastructure logicielle opérationnelle, modulaire et évolutive, établit les liens entre les outils (briques logicielles) et le modèle de données INSPIRE étendu introduit précédemment. En veillant à la qualité des données et métadonnées, elle garantit une restitution riche et précise de l'information statistique. Le respect des normes et l'architecture modulaire confèrent au système une capacité d'adaptation à divers domaines.

La Figure 43 donne une vue d'ensemble de l'infrastructure logicielle STedi. Le composant SUNI (pour *Statistical Units Nomenclature Integrator*, numéroté 1 dans la figure) permet d'intégrer des nomenclatures territoriales dans une base de données relationnelle reposant sur le modèle de données territoriales évolutives initialement décrit dans (Plumejeaud, 2011). Pour opérationnaliser le recours au modèle de métadonnées socio-économiques évoqué plus haut et garantir le respect des directives associées, un outil interactif de suivi des dépôts de jeux de données en ligne a été créé (Bernard et al., 2014; Ysebaert et al., 2014). Cet outil, appelé le *Checking tool* (Figure 43, numéro 2), aide le producteur de données à améliorer la qualité de ses dépôts par le biais de rapports d'expertise. Ces rapports deviennent les documents de référence pour l'évaluation et l'amélioration de la qualité des jeux de données avant leur intégration en base de données. Le *Checking tool* inclut d'autres composants tels que le composant *Dataset Check*, chargé de vérifier la conformité syntaxique d'un jeu de données. Le composant *Dataset Integrator* permet d'intégrer un jeu de données dans la base de données (Figure 43, numéro 3), il convertit des éléments du jeu de données, délivré au format XLS, en objets Java manipulables. Le composant *Data Access* offre des fonctionnalités de recherche pour accéder aux données dans la base. Ce composant peut être utilisé par un autre composant du type *web application* pour accéder de manière spécifique aux données (Figure 43, numéro 4). Il intervient aussi

lors de la restitution de données via des web services OGC (*Open Geospatial Consortium*) du type WFS (*Web Feature Service*) ou WMS (*Web Map Service*) (Figure 43, numéro 5).

Ainsi, l'infrastructure implémentée dans le cadre du programme ESPON est une instance de l'infrastructure logicielle STedi. Dans cette instance, le composant *web application* notamment répond aux besoins spécifiques des commanditaires du programme ESPON, notamment en ce qui concerne l'interface de recherche qui permet d'interroger et de consulter les données, pour lesquelles sont générées des fiches de métadonnées précises. Cette application web est accessible à l'adresse <http://database.espon.eu/db2/search>.

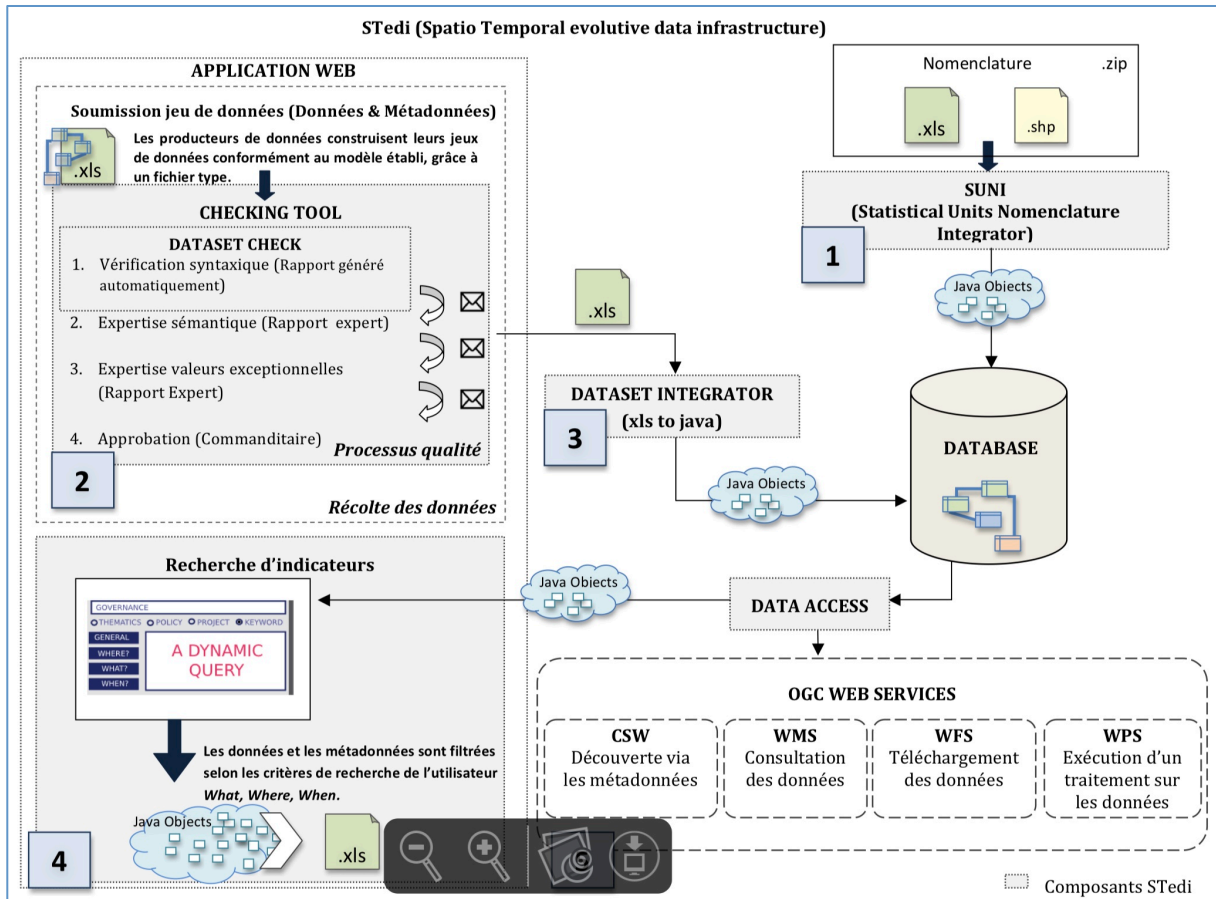


Figure 43 - Vue générale de l'infrastructure logicielle STedi

Source : (Bernard et al., 2014)

Annexe 2 – Projet ANR CHOUCAS

Sur le territoire français, plus de 15000 demandes de secours en montagne sont recensées chaque année dont la grande majorité sont gérées par les Pelotons de Gendarmerie de Haute-Montagne (PGHM). Parvenir à localiser rapidement les victimes lorsqu'une alerte est donnée reste parfois difficile pour les équipes de secours, alors que réduire le temps d'identification de la zone où se trouve la victime est souvent déterminant pour une issue favorable du sauvetage. Parmi les outils récemment développés dont disposent les secouristes, l'outil GendLoc¹³⁰ est un système de géolocalisation qui exploite la puce GPS des *smartphones*. Ce système est fonctionnel dès lors qu'un réseau est disponible, sans qu'aucune application n'ait été téléchargée au préalable, et repose sur le principe de localisation « consentie », c'est-à-dire autorisée par l'appelant. En effet, suite à la réception d'un appel, un agent du centre de secours peut envoyer sur le téléphone de l'appelant un SMS contenant un lien HTML. Lorsque l'appelant clique sur ce lien, les coordonnées de sa position sont renvoyées et l'agent peut visualiser cette position sur son écran. Les avantages du système sont un gain de temps important dans la localisation des victimes et dans l'acheminement des secours, notamment du fait de la précision des informations de localisation obtenues. En outre, dans de nombreux cas, il permet de proposer une assistance à distance, adaptée à une situation non critique (comme dans le cas de randonneurs égarés), contribuant ainsi à réaliser des économies non négligeables en termes de ressources (humaines et autres : véhicules, hélicoptères, etc.) et à maintenir un niveau de réactivité élevé en vue d'autres situations de danger qui pourraient survenir.

GendLoc est un outil pertinent, adopté depuis 2015 par tous les PGHM de France (y compris hors métropole) et d'autres acteurs du secours : CORG (Centre d'Opérations et de Renseignement de la Gendarmerie), CRS (Compagnies Républicaines de Sécurité), unités de secours en mer, etc. (Muscat, 2015). Il s'est notamment révélé particulièrement intéressant en forêt, de nuit ou par mauvais temps, lorsque la victime n'est pas visible du ciel ou encore a des difficultés à exprimer sa localisation. Néanmoins, il existe plusieurs situations pour lesquelles le recours à ce système n'est pas envisageable. On peut citer à titre d'exemple les cas suivants :

- lorsque la personne à secourir n'a pas de *smartphone* mais un portable d'ancienne génération ne permettant pas l'utilisation du système,
- lorsque l'état – physique ou de conscience – de l'appelant ne lui permet pas de réaliser la manipulation requise pour autoriser la localisation,
- lorsque la couverture réseau ou la charge de batterie sont défaillantes et rendent chaotique l'activation du processus de localisation,
- lorsque l'appelant est un tiers qui n'est pas à proximité de la victime mais l'a vue, par exemple, tomber à distance¹³¹,
- ou encore lorsque l'appelant est un tiers qui s'inquiète de ne pas voir rentrer un proche d'une sortie en montagne.

Pour assurer sa mission de secours dans ce type de situations, disposer d'un autre système d'aide à la localisation de victimes est un enjeu important pour un acteur tel qu'un PGHM.

Le projet ANR CHOUCAS¹³² a l'ambition de contribuer à la mise au point d'un tel outil (Olteanu-Raimond et al., 2017). Ce projet implique des chercheurs¹³³ en raisonnement spatial, en gestion de données et de connaissances, en extraction d'information, en géovisualisation et, les

¹³⁰ <https://fr.wikipedia.org/wiki/GendLoc> - GendLoc a été mis au point par l'adjudant Olivier Favre, gendarme du PGHM du Versoud en Isère.

¹³¹ des cordées qui dévissent, des avalanches, ou encore des accidents de parapente ou de vol en *wingsuit* (combinaison ailée) peuvent correspondre à ces situations, avec des distances de l'ordre potentiellement du kilomètre.

¹³² <http://choucas.ign.fr/> - *Choucas* est l'indicatif radio des hélicoptères de la gendarmerie, utilisés, entre autres, pour le secours en montagne.

¹³³ Membres du COGIT de l'IGN, de l'équipe MOVIES du Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour et de l'équipe STEAMER du Laboratoire d'Informatique de Grenoble.

professionnels du secours en montagne à travers le PGHM du Versoud¹³⁴. Plusieurs objectifs scientifiques sont poursuivis dans ce contexte par les partenaires du projet, comme la constitution et l'enrichissement de données géographiques issues de sources hétérogènes (Medad et al., 2018) ou la proposition d'approches de raisonnement spatial flou (Bunel et al., 2018). Le travail mené au sein de l'équipe STeamer porte plus particulièrement sur la modélisation des connaissances métier et la façon dont elles sont convoquées dans le processus de raisonnement du secouriste, avec comme objectif de proposer un environnement de géovisualisation contribuant à améliorer ce processus, dont la finalité est la mise en exergue de la localisation probable des victimes. Nos travaux à venir sur cette thématique sont décrits dans la section 5.2.

¹³⁴ Dont, en particulier, l'adjudant O. Favre, inventeur de GendLoc.

Références bibliographiques

Abi-Zeid, I., Nilo, O., Schwartz, S., Morin, M., 2010. Towards a knowledge-based system prototype for aeronautical Search and Rescue operations, in: 2010 13th International Conference on Information Fusion. Presented at the 2010 13th International Conference on Information Fusion (FUSION 2010), IEEE, Edinburgh, pp. 1–8. <https://doi.org/10.1109/ICIF.2010.5712112>

Abler, R., Adams, J.S., Gould, P., 1971. Spatial organization: the geographer's view of the world. Prentice-Hall, Englewood Cliffs, NJ.

Al-Bakri, M., Atencia, M., David, J., Lalande, S., Rousset, M.-C., 2016. Uncertainty-Sensitive Reasoning for Inferring sameAs Facts in Linked Data. *Front. Artif. Intell. Appl.* 698–706. <https://doi.org/10.3233/978-1-61499-672-9-698>

Allen, J.F., 1984. Towards a general theory of action and time. *Artif. Intell.* 23, 123–154. [https://doi.org/10.1016/0004-3702\(84\)90008-0](https://doi.org/10.1016/0004-3702(84)90008-0)

Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A., Moelans, B., Palma, A.T., 2007a. Towards Semantic Trajectory Knowledge Discovery (Technical Report). Hasselt University.

Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A., 2007b. A Model for Enriching Trajectories with Semantic Geographical Information, in: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, GIS '07. ACM, New York, NY, USA, pp. 22:1–22:8. <https://doi.org/10.1145/1341012.1341041>

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M. -J., MacEachren, A., Wrobel, S., 2007. Geovisual analytics for spatial decision support: Setting the research agenda. *Int. J. Geogr. Inf. Sci.* 21, 839–857. <https://doi.org/10.1080/13658810701349011>

Arnaud, A., Davoine, P.A., 2011. Approche cartographique et géovisualisation pour la représentation de l'incertitude. Application à l'information dédiée aux risques naturels. *Rev. Int. Géomat.* 21, 205–224. <https://doi.org/10.3166/geo.21.205-224>

Atkinson, R., 2017. QB4ST: RDF Data Cube extensions for spatio-temporal components [WWW Document]. URL <https://www.w3.org/TR/qb4st/> (accessed 3.1.17).

Authier, J.Y., Bidet, J., Collet, A., Gilbert, P., Steinmetz, H., 2010. État des lieux sur les trajectoires résidentielles (PUCA).

Bachimont, B., Isaac, A., Troncy, R., 2002. Semantic commitment for designing ontologies: a proposal, in: International Conference on Knowledge Engineering and Knowledge Management. Springer, pp. 114–121.

Baglioni, M., Macedo, J., Renso, C., Wachowicz, M., 2008. An Ontology-Based Approach for the Semantic Modelling and Reasoning on Trajectories, in: Song, I.-Y., Piattini, M., Chen, Y.-P.P., Hartmann, S., Grandi, F., Trujillo, J., Opdahl, A.L., Ferri, F., Grifoni, P., Caschera, M.C., Rolland, C., Woo, C., Salinesi, C., Zimányi, E., Claramunt, C., Frasincar, F.,

Houben, G.-J., Thiran, P. (Eds.), *Advances in Conceptual Modeling – Challenges and Opportunities*. Springer Berlin Heidelberg, pp. 344–353.

Banaee, H., 2018. *From Numerical Sensor Data to Semantic Representations : A Data-driven Approach for Generating Linguistic Descriptions* (PhD Thesis). Örebro University, School of Science and Technology.

Beard, K., Mackaness, W., 1993. Visual Access to Data Quality in Geographic Information Systems. *Cartogr. Int. J. Geogr. Inf. Geovisualization* 30, 37–45. <https://doi.org/10.3138/C205-5885-23M7-0664>

Bénaben, F., Hanachi, C., Lauras, M., Couget, P., 2008. A metamodel and its ontology to guide crisis characterization and its collaborative management.

Bernard, C., Le Rubrus, B., Villanova-Oliver, M., Gensel, J., Ysebaert, R., Salmon, I., 2014. STedi: une infrastructure logicielle pour renforcer la qualité des données territoriales statistiques, in: *SAGEO Spatial Analysis and GEomatics*.

Bernard, C., Plumejeaud, C., Villanova-Oliver, M., Gensel, J., Dao, H., 2018a. An Ontology-based Algorithm for Managing the Evolution of Multi-Level Territorial Partitions, in: *26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '18)*. Seattle, Washington, USA.

Bernard, C., Villanova-Oliver, M., Gensel, J., Dao, H., 2018b. Modeling changes in territorial partitions overtime: ontologies TSN and TSN-change, in: *Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC 2018, Pau, France, April 09-13, 2018*. pp. 866–875. <https://doi.org/10.1145/3167132.3167227>

Bernard, C., Villanova-Oliver, M., Gensel, J., Dao, H., 2017a. TSN et TSN-change : ontologies pour représenter l'évolution des découpages territoriaux statistiques, in: *SAGEO 2017 - Spatial Analysis and GEomatics*. GdR Magis CNRS, Rouen, France.

Bernard, C., Villanova-Oliver, M., Gensel, J., Le Rubrus, B., 2017b. Spatio-Temporal evolutive Data Infrastructure: a Spatial Data Infrastructure for managing data flows of Territorial Statistical Information. *Int. J. Digit. Earth* 10, 257–283. <https://doi.org/10.1080/17538947.2016.1222003>

Berners-Lee, T., Hendler, J., Lassila, O., others, 2001. The semantic web. *Sci. Am.* 284, 28–37.

Bilasco, I.M., 2007. *Une approche sémantique pour la réutilisation et l'adaptation de données 3D* (Theses). Université Joseph-Fourier - Grenoble I.

Bittner, T., Smith, B., 2003. A Theory of Granular Partitions, in: Worboys, M., Duckham, M., Goodchild, M. (Eds.), *Foundations of Geographic Information Science*. CRC Press, pp. 117–149. <https://doi.org/10.1201/9780203009543.ch7>

Blaise, J.-Y., Davoine, P.-A., Kaddouri, L., Mathian, H., Saint-Marc, C., 2015. Géovisualisation des temporalités des territoires : analyse comparative de la diversité des productions, in: *Douzième Rencontres de Théo Quant*. Besançon, France.

Blomqvist, E., Sandkuhl, K., 2005. Patterns in Ontology Engineering – Classification of Ontology Patterns, in: *ICEIS 2005 : Proceedings of the Seventh International Conference on Enterprise Information Systems, Miami, USA, May 25-28, 2005*. School of Engineering, Jönköping University, JTH. Research area Information Engineering, pp. 4–.

Bogorny, V., Renso, C., de Aquino, A.R., de Lucca Siqueira, F., Alvares, L.O., 2014. CONSTAnT - A Conceptual Data Model for Semantic Trajectories of Moving Objects: CONSTAnT - A Conceptual Data Model for Semantic Trajectories. *Trans. GIS* 18, 66–88. <https://doi.org/10.1111/tgis.12011>

Bonvalet, C., Bringé, A., 2012. Les trajectoires géographiques des Franciliens depuis leur départ de chez les parents », in: Bonvalet, C., Lelièvre, E. (Eds.), *De La Famille à l'entourage*.

INED, Paris, pp. 177–202.

Boudou, M., 2015. Approche multidisciplinaire pour la caractérisation d'inondations remarquables : enseignements tirés de neuf évènements en France (1910-2010) (Theses). Université Paul Valéry - Montpellier III.

Bratman, M., 1999. Intention, plans, and practical reason, David Hume series. Center for the Study of Language and Information, Stanford, Calif.

Bunel, M., Olteanu-Raimond, A.-M., Duchêne, C., 2018. Référencement spatial indirect : modélisation à base de relations et d'objets spatiaux vagues., in: Sageo'2018. Presented at the Spatial Analysis and GEomatics, Montpellier, France.

Burkhardt, J., Donikian, S., Duthen, Y., Héguy, O., Lourdeau, D., Morineau, T., Parenthoën, M., Sanza, C., Tisseau, J., 2003. Thème 5 : Comportements humains individuels et collectifs. (Rapport Action Spécifique du CNRS), Humain virtuel : vers un humain synthétique temps-réel aussi vrai que nature.

Buttenfield, B.P., 1993. Representing Data Quality. *Cartogr. Int. J. Geogr. Inf. Geovisualization* 30, 1–7. <https://doi.org/10.3138/232H-6766-3723-5114>

Caglioni, M., Fusco, G., 2014. Formal Ontologies and Uncertainty. In *Geographical Knowledge. Tema J. Land Use Mobil. Environ. 2014 INPUT 2014 - Smart City Plan. Energy Transp. Sustain. Urban Syst.* <https://doi.org/10.6092/1970-9870/2530>

Canellas, M.C., Feigh, K.M., Chua, Z.K., 2015. Accuracy and Effort of Decision-Making Strategies With Incomplete Information: Implications for Decision Support System Design. *IEEE Trans. Hum.-Mach. Syst.* 45, 686–701. <https://doi.org/10.1109/THMS.2015.2420575>

Caquard, S., Cartwright, W., 2014. Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping. *Cartogr. J.* 51, 101–106. <https://doi.org/10.1179/0008704114Z.000000000130>

Card, S., 2008. Information Visualization, in: *The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Jacko, Julie and Sears, Andrew, pp. 509–543.

Card, S.K., Mackinlay, J.D., Shneiderman, B., 1999. Readings in information visualization: using vision to think, *The Morgan Kaufmann series in interactive technologies*. Morgan Kaufmann Publishers, San Francisco, Calif.

Carrillo Ramos, A.C., 2007. Agents ubiquitaires pour un accès adapté aux systèmes d'information : le framework PUMAS.

Chapurlat, V., Aloui, S., 2007. Une démarche intégrée pour l'ingénierie de système complexe face aux risques, in: *7ème Congrès International de Génie Industriel*. Trois rivières, Canada, p. 11p.

Chazelle, B., Lambert, L., Pams Capoccioni, C., 2014. La vulnérabilité des infrastructures ferroviaires face aux évènements hydrologiques extrêmes. État des connaissances et gestion du risque. *Houille Blanche* 48–54. <https://doi.org/10.1051/lhb/2014016>

Chen, C., 2005. Top 10 Unsolved Information Visualization Problems. *IEEE Comput. Graph. Appl.* 25, 12–16. <https://doi.org/10.1109/MCG.2005.91>

Christophe, S., 2017. Personnalisation des représentations cartographiques en géovisualisation : couleurs et styles (HDR). Université Paris Est.

Clark, W.A.V., Davies Withers, S., 2007. Family migration and mobility sequences in the United States: Spatial mobility in the context of the life course. *Demogr. Res.* 17, 591–622. <https://doi.org/10.4054/DemRes.2007.17.20>

Codet de Boisse, A., 2013. Aide à la décision exploitant de la connaissance générale et contextuelle : application à la maintenance d'hélicoptère.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S.,

Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K., 2012. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semant. Sci. Serv. Agents World Wide Web* 17, 25–32. <https://doi.org/10.1016/j.websem.2012.05.003>

Conrad, S.H., LeClaire, R.J., O'Reilly, G.P., Uzunalioglu, H., 2006. Critical national infrastructure reliability modeling and analysis. *Bell Labs Tech. J.* 11, 57–71. <https://doi.org/10.1002/bltj.20178>

Conradi, R., Westfechtel, B., 1998. Version models for software configuration management. *ACM Comput. Surv.* CSUR 30, 232–282.

Correndo, G., Granzotto, A., Salvadores, M., Hall, W., Shadbolt, N., 2010. A linked data representation of the nomenclature of territorial units for statistics.

Correndo, G., Shadbolt, N., 2013. Linked nomenclature of territorial units for statistics. *Semantic Web* 4, 251–256.

Couix, S., Burkhardt, J.-M., Donikian, S., Rouillé, M., 2005. Etat de l'art sur les modèles de comportements et les modèles de tâches pour la génération d'opérateurs virtuels au travail (Livrable No. 2.1.0), *Projet RNTL PerfRV2*.

Coulter, R., van Ham, M., 2013. Following People Through Time: An Analysis of Individual Residential Mobility Biographies. *Hous. Stud.* 28, 1037–1055. <https://doi.org/10.1080/02673037.2013.783903>

Cygniak, R., Reynolds, D., 2014. The RDF Data Cube Vocabulary [WWW Document]. URL <https://www.w3.org/TR/vocab-data-cube/> (accessed 12.11.17).

Davoine, P.-A., 2014. Contributions géomatiques pour la Gestion des Risques Naturels : modélisation, géovisualisation, acquisition. Grenoble INP, Grenoble (France).

Davoine, P.-A., Saint-Marc, C., Di Muro, A., Staudacher, T., Villanova-Oliver, M., Boissier, P., 2012. Constructions cartographiques pour la caractérisation de l'aléa volcanologique du Piton de la Fournaise : Cartographie de séries de données chrono-spatiales, in: *Actes de La Conférence SAGEO, 7-9 Novembre 2012*. Liège, Belgium, p. (16 pages).

Del Mondo, G., 2011. UN MODELE DE GRAPHE SPATIO-TEMPOREL POUR REPRESENTER L'EVOLUTION D'ENTITES GEOGRAPHIQUES (phdthesis). Université de Bretagne occidentale - Brest.

Dibiase, D., MacEachren, A.M., Krygier, J.B., Reeves, C., 1992. Animation and the Role of Map Design in Scientific Visualization. *Cartogr. Geogr. Inf. Syst.* 19, 201–214. <https://doi.org/10.1559/152304092783721295>

Dubos-Paillard, E., Provitolo, D., 2012. Analysis of the 2011 Japan's triple disaster from a conceptual ontology of the domain of risk and disaster, in: *Complex Systems Design and Management*. France, pp. 1–14.

Ermine, J.-L., 2003. *La gestion des connaissances*. Hermes Science : Lavoisier, Paris.

Fabrikant, S., 2001. Building Task-Ontologies for GeoVisualization. Presented at the Pre-Conference Workshop on Geovisualization on the Web, Beijing, China.

Ferrero, C.A., Alvares, L.O., Bogorny, V., 2016. Multiple aspect trajectory data analysis: research challenges and opportunities, in: Campelo, C.E.C., Namikawa, L.M. (Eds.), *XVII Brazilian Symposium on Geoinformatics - GeoInfo 2016*, Campos Do Jordão, SP, Brazil, November 27-30, 2016. MCTIC/INPE, pp. 56–67.

Flowerdew, R., 1991. Data integration: statistical methods for transferring data between zonal systems. *Handl. Geogr. Inf. Methodol. Potential Appl.* 39–54.

Fusco, G., 2012. Démarche géo-prospective et modélisation causale probabiliste. *Cybergeo*. <https://doi.org/10.4000/cybergeo.25423>

- Gautier, J., 2018. GRAPHIST : une approche exploratoire pour l'identification des dynamiques des phénomènes spatio-temporels. Université Grenoble-Alpes, Grenoble, France.
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69, 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Goodchild, M.F., Lam, N.S.N., Geography, U. of W.O.D. of, 1980. Areal interpolation: a variant of the traditional spatial problem. London, Ont.: Department of Geography, University of Western Ontario.
- Grafmeyer, Y., Authier, J.Y., 2011. *Sociologie urbaine*. Armand Colin.
- Griethe, H., Schumann, H., 2005. Visualizing uncertainty for improved decision making.
- Guarino, N., Welty, C., 2000. A formal ontology of properties, in: *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*. Springer, pp. 97–112.
- Güting, R.H., Schneider, M., 2005. *Moving objects databases*, Morgan Kaufmann series in data management systems. Morgan Kaufmann, Amsterdam : Boston.
- Hägerstrand, T., 1970. WHAT ABOUT PEOPLE IN REGIONAL SCIENCE? *Pap. Reg. Sci.* 24, 7–24. <https://doi.org/10.1111/j.1435-5597.1970.tb01464.x>
- Hajji, H., 2005. *Gestion des risques naturels : une approche fondée sur l'intégration des données* (PhD Thesis).
- Haugeland, J., 1985. *Artificial intelligence: the very idea*. MIT Press, Cambridge, Mass.
- Heinrich, H.W., 1931. *Industrial Accident Prevention: A Scientific Approach*, McGraw-Hill insurance series. McGraw-Hill book Company, Incorporated.
- Holmqvist, K., Andersson, R., 2017. *Eye tracking: a comprehensive guide to methods, paradigms, and measures*.
- Hombiat, A., 2017. OF4OSM : un méta-modèle pour structurer la folksonomie d'OpenStreetMap en une nouvelle ontologie. Grenoble Alpes.
- Hornsby, K., Egenhofer, M.J., 2002. Modeling Moving Objects over Multiple Granularities. *Ann. Math. Artif. Intell.* 36, 177–194. <https://doi.org/10.1023/A:1015812206586>
- Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M., Kolas, D., 2013. A Geo-ontology Design Pattern for Semantic Trajectories, in: Tenbrink, T., Stell, J., Galton, A., Wood, Z. (Eds.), *Spatial Information Theory*. Springer International Publishing, Cham, pp. 438–456. https://doi.org/10.1007/978-3-319-01790-7_24
- Janowicz, K., Hitzler, P., 2013. Thoughts on the complex relation between linked data, semantic annotations, and ontologies, in: *Proceedings of the Sixth International Workshop on Exploiting Semantic Annotations in Information Retrieval - ESAIR '13*. Presented at the the sixth international workshop, ACM Press, San Francisco, California, USA, pp. 41–44. <https://doi.org/10.1145/2513204.2513218>
- Jensen, C.S., Dyreson, C.E., Böhlen, M., Clifford, J., Elmasri, R., Gadia, S.K., Grandi, F., Hayes, P., Jajodia, S., Käfer, W., Kline, N., Lorentzos, N., Mitsopoulos, Y., Montanari, A., Nonen, D., Peressi, E., Pernici, B., Roddick, J.F., Sarda, N.L., Scalas, M.R., Segev, A., Snodgrass, R.T., Soo, M.D., Tansel, A., Tiberio, P., Wiederhold, G., 1998. The consensus glossary of temporal database concepts — February 1998 version, in: Etzion, O., Jajodia, S., Sripada, S. (Eds.), *Temporal Databases: Research and Practice*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 367–405. <https://doi.org/10.1007/BFb0053710>
- Jin, M., Claramunt, C., 2018. A Semantic Model for Human Mobility in an Urban Region. *J. Data Semant.* 7, 171–187. <https://doi.org/10.1007/s13740-018-0092-4>
- Kauppinen, T., Hyvönen, E., 2007. Modeling and reasoning about changes in ontology time series, in: *Ontologies*. Springer, pp. 319–338.

Kauppinen, T., Väätäinen, J., Hyvönen, E., 2008. Creating and using geospatial ontology time series in a semantic cultural heritage portal. Springer.

Kinkeldey, C., MacEachren, A.M., Riveiro, M., Schiewe, J., 2017. Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartogr. Geogr. Inf. Sci.* 44, 1–21. <https://doi.org/10.1080/15230406.2015.1089792>

Kinkeldey, C., MacEachren, A.M., Schiewe, J., 2014. How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. *Cartogr. J.* 51, 372–386. <https://doi.org/10.1179/1743277414Y.0000000099>

Kirsch Pinheiro, M., 2006. Adaptation contextuelle et personnalisée de l'information de conscience de groupe au sein des systèmes d'information coopératifs.

Klein, G.A., 1998. Sources of power: how people make decisions. MIT Press, Cambridge, Mass.

Klein, O., 2007. Modélisation et représentations spatio-temporelles des déplacements quotidiens urbains. Université Louis Pasteur.

Klien, E., Probst, F., 2005. Requirements for geospatial ontology engineering, in: 8th Conference on Geographic Information Science (AGILE 2005). pp. 251–260.

Kruchten, P., Woo, C., Monu, K., Sotoodeh, M., 2007. A human-centered conceptual model of disasters affecting critical infrastructures.

Lang, M., Coeur, D., 2014. Les inondations remarquables en France: inventaire 2011 pour la directive Inondation. Ed. Quae, Versailles.

Le Fur, A., 2007. Pratiques de la cartographie, 2ème édition. ed. Armand Colin.

Le Moigne, J.-L., 1999. La modélisation des systèmes complexes, Hors Collection. Dunod.

Le Pors, T., 2010. Simulation cognitive de la prise de décision d'experts ; application au trafic maritime (Theses). Université de Bretagne Sud.

Lelévrier, C., 2007. Mobilités et trajectoires résidentielles des ménages dans trois opérations de rénovation urbaine en Île-de-France, (PUCA).

Levenshtein, V.I., 1966. Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Dokl.* 10, 707–710.

Lodemann, M., Luttenberger, N., 2010. Ontology-based Railway Infrastructure Verification - Planning Benefits.

Lopez-Velasco, C., 2008. Sélection et composition de services Web pour la génération d'applications adaptées au contexte d'utilisation (Theses). Université Joseph-Fourier - Grenoble I.

Lu, J., Roberts, C., Lang, K., Stirling, A., Madelin, K., 2006. The Application of Semantic Web Technologies for Railway Decision Support, in: Intelligent Decision-Making Support Systems. Springer London, London, pp. 321–337. https://doi.org/10.1007/1-84628-231-4_17

MacEachren, A.M., Gahegan, M., Pike, W., Brewer, I., Cai, G., Lengerich, E., Hardistry, F., 2004. Geovisualization for knowledge construction and decision support. *IEEE Comput. Graph. Appl.* 24, 13–17. <https://doi.org/10.1109/MCG.2004.1255801>

MacEachren, A.M., Kraak, M.-J., 1997. Exploratory cartographic visualization: Advancing the agenda. *Comput. Geosci.* 23, 335–343. [https://doi.org/10.1016/S0098-3004\(97\)00018-6](https://doi.org/10.1016/S0098-3004(97)00018-6)

MacEachren, A.M., Roth, R.E., O'Brien, J., Li, B., Swingley, D., Gahegan, M., 2012.

Visual Semiotics & Uncertainty Visualization: An Empirical Study. *IEEE Trans. Vis. Comput. Graph.* 18, 2496–2505. <https://doi.org/10.1109/TVCG.2012.279>

McCormick, B., DeFanti, T., Brown, M., 1987. Visualization in Scientific Computing-A Synopsis. *IEEE Comput. Graph. Appl.* 7, 61–70. <https://doi.org/10.1109/MCG.1987.277014>

Medad, A., Gaio, M., Le Nir, Y., Mustière, S., 2018. Appariement automatique de données hétérogènes: textes, traces GPS et ressources géographiques, in: *Sagéo'2018*. Montpellier, France.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. <https://doi.org/10.1037/h0043158>

Miron, A.D., 2009. Découverte d'associations sémantique pour le Web sémantique géospatial : le framework ONTOAST.

Morris, C., Easton, J., Roberts, C., 2015. Position Paper: Ontology in the Rail Domain - The Railway Core Ontologies: SCITEPRESS - Science and Technology Publications, pp. 285–290. <https://doi.org/10.5220/0005613702850290>

Mulder, C.H., Hooimeijer, P., 1999. Residential Relocations in the Life Course, in: Wissen, L.J.G., Dykstra, P.A. (Eds.), *Population Issues*. Springer Netherlands, Dordrecht, pp. 159–186. https://doi.org/10.1007/978-94-011-4389-9_6

Mulder, C.H., Wagner, M., 1993. Migration and marriage in the life course: a method for studying synchronized events. *Eur. J. Popul.* 9, 55–76. <https://doi.org/10.1007/BF01267901>

Muscat, A., 2015. GendLoc : une géolocalisation simple et rapide. *Gend'Info* 39.

Noel, D., Villanova-Oliver, M., Gensel, J., Le Quéau, P., 2017. Design Patterns for Modelling Life Trajectories in the Semantic Web, in: David Brosset, T.W., Christophe Claramunt, Xiang Li (Ed.), *International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2017)*, Web and Wireless Geographical Information Systems 15th International Symposium, W2GIS 2017, Shanghai, China, May 8-9, 2017, Proceedings. Springer, Shanghai, China, pp. 51–65. https://doi.org/10.1007/978-3-319-55998-8_4

Noël, D., Villanova-Oliver, M., Gensel, J., Le Quéau, P., 2017. Design patterns for modelling life trajectories in the semantic web, in: *15th International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2017)*, 15th International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2017). Shanghai, China.

Noël, D., Villanova-Oliver, M., Gensel, J., Le Quéau, P., 2016. Modélisation de trajectoires sémantiques intégrant perspectives multiples et facteurs explicatifs: Application aux trajectoires de vie. *Rev. Int. Géomat.* 26, 491–510. <https://doi.org/10.3166/rig.2016.00010>

Noël, D., Villanova-Oliver, M., Gensel, J., Le Quéau, P., 2015. Modeling semantic trajectories including multiple viewpoints and explanatory factors: application to life trajectories, in: *ACM SIGSPATIAL International Workshop on Smart Cities and Urban Analytics (UrbanGIS) 2015*. ACM SIGSPATIAL, Seattle, United States.

Noy, N.F., Klein, M., 2004. Ontology Evolution: Not the Same as Schema Evolution. *Knowl. Inf. Syst.* 6, 428–440. <https://doi.org/10.1007/s10115-003-0137-2>

Olteanu-Raimond, A.-M., Davoine, P.-A., Gaio, M., gouarderes, eric, Van Damme, M.-D., Villanova-Oliver, M., Brasebin, M., Domingues, C., Duchêne, C., Favre, O., Mustière, S., Devin, F., Le Nir, Y., Moncla, L., Bouveret, S., Genoud, P., Gensel, J., Ziebelin, D., 2017. *Projet CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne*, in: *Spatial Analysis and GEOMatics 2017*. INSA de rouen, Rouen, France.

Openshaw, S., Taylor, P.J., 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Stat. Appl. Spat. Sci.* 21, 127–144.

Parent, C., Pelekis, N., Theodoridis, Y., Yan, Z., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., 2013. Semantic trajectories modeling and analysis. *ACM Comput. Surv.* 45, 1–32. <https://doi.org/10.1145/2501654.2501656>

Payne, J.W., Bettman, J.R., Johnson, E.J., 1993. *The adaptive decision maker*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139173933>

Plumejeaud, C., 2011. *Modèles et méthodes pour l'information spatio-temporelle évolutive*. Grenoble.

Plumejeaud, C., Azzi, D., Villanova-Oliver, M., Gensel, J., 2011a. Toward an interactive system for checking spatio-temporal data quality, in: *Proceedings of the 1st Conference on Spatial Statistics 2011, Mapping Global Change*. Unknown.

Plumejeaud, C., Gensel, J., Villanova-Oliver, M., 2010. Opérationnalisation d'un profil ISO 19115 pour des métadonnées socio-économiques, in: *INFORSID*. Unknown, pp. 25–41.

Plumejeaud, C., Gensel, J., Villanova-Oliver, M., Ben Rebah, M., Vergnaud, G., 2009. Modélisation de hiérarchies territoriales multiples - Vers la gestion d'informations spatio-temporelles évolutives, in: *Colloque International de Géomatique et d'Analyse Spatiale - SAGEO 2009*. Paris, France.

Plumejeaud, C., Rebah, M.B., Gensel, J., Villanova-Oliver, M., 2011b. Construction d'une généalogie des territoires pour la représentation du changement. *Cybergeog.* <https://doi.org/10.4000/cybergeog.24851>

Plumejeaud, C., Villanova-Oliver, M., 2012. QualESTIM: Interactive quality assessment of socio-economic data using outlier detection, in: Gensel, J., Josselin, D., Vandenbroucke, D. (Eds.), *Bringing the Geographic Information Sciences, Proceedings of International AGILE'2012 Conference*. Unknown, pp. 143–160.

Poole, A., Ball, L.J., 2005. Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future. *Encycl. Hum.-Comput. Interact.*

Provitolo, D., Dubos-Paillard, E., Müller, J.-P., 2009. Vers une ontologie des risques et des catastrophes: le modèle conceptuel, in: *Ontologie et Dynamique Des Systèmes Complexes, Perspectives Interdisciplinaires*. Rochebrune (Megève), France, pp. 1–16.

Raja Mohan, A., Arumugam, G., 2011. Constructing Railway Ontology using Web Ontology Language and Semantic Web Rule Language. *International Journal of Computer Technology and Applications* 2, 314-321.

Robinson, A., 2017. Geovisual Analytics. *Geogr. Inf. Sci. Technol. Body Knowl.* 2017. <https://doi.org/10.22224/gistbok/2017.3.6>

Robinson, A.C., Demšar, U., Moore, A.B., Buckley, A., Jiang, B., Field, K., Kraak, M.-J., Camboim, S.P., Sluter, C.R., 2017. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *Int. J. Cartogr.* 3, 32–60. <https://doi.org/10.1080/23729333.2016.1278151>

Saad, I., Grundstein, M., Rosenthal-Sabroux, C., 2009. Une méthode d'aide à l'identification des connaissances cruciales pour l'entreprise. *Systèmes Inf. Manag.* 14, 43. <https://doi.org/10.3917/sim.093.0043>

Saint-Marc, C., 2017. Formalisation et géovisualisation d'événements historiques issus de risques naturels pour la compréhension des dynamiques spatiales: Application aux inondations ayant touché le système ferroviaire français. *Université Grenoble-Alpes*.

Saint-Marc, C., Villanova-Oliver, M., Davoine, P.-A., Pams Capoccioni, C., Chenier, D., 2016. Representation and Visualization of Imperfect Geohistorical Data About Natural Risks: A Qualitative Classification and Its Experimental Assessment, in: Sarjakoski, T., Santos, M.Y., Sarjakoski, L.T. (Eds.), *19th AGILE Conference on Geographic Information Science*,

Lecture Notes in Geoinformation and Cartography. Springer, Helsinki, Finland. https://doi.org/10.1007/978-3-319-33783-8_14

Santos, R., Murrieta-Flores, P., Martins, B., 2018. Learning to combine multiple string similarity metrics for effective toponym matching. *Int. J. Digit. Earth* 11, 913–938. <https://doi.org/10.1080/17538947.2017.1371253>

Scheuer, S., Haase, D., Meyer, V., 2013. Towards a flood risk assessment ontology – Knowledge integration into a multi-criteria risk assessment approach. *Comput. Environ. Urban Syst.* 37, 82–94. <https://doi.org/10.1016/j.compenvurbsys.2012.07.007>

Silva, R.A., Pires, J.M., Santos, M.Y., 2017. When Granules Are not Enough in a Theory of Granularities, in: Bregt, A., Sarjakoski, T., van Lammeren, R., Rip, F. (Eds.), *Societal Geo-Innovation*. Springer International Publishing, Cham, pp. 279–296. https://doi.org/10.1007/978-3-319-56759-4_16

Slocum, T.A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D.R., Fuhrmann, S., Hedley, N.R., 2001. Cognitive and Usability Issues in Geovisualization. *Cartogr. Geogr. Inf. Sci.* 28, 61–75. <https://doi.org/10.1559/152304001782173998>

Spaccapietra, S., Parent, C., Damiani, M.L., de Macedo, J.A., Porto, F., Vangenot, C., 2008. A conceptual view on trajectories. *Data Knowl. Eng.* 65, 126–146. <https://doi.org/10.1016/j.datak.2007.10.008>

Sreeves, M., 2018. *Managing Uncertainty in Mountain Search and Rescue (Master 2 Recherche MOSIG)*. Grenoble INP.

Stefanidis, K., Chrysakis, I., Flouris, G., 2014. On Designing Archiving Policies for Evolving RDF Datasets on the Web, in: Yu, E., Dobbie, G., Jarke, M., Puro, S. (Eds.), *Conceptual Modeling: 33rd International Conference, ER 2014, Atlanta, GA, USA, October 27-29, 2014. Proceedings*. Springer International Publishing, Cham, pp. 43–56. https://doi.org/10.1007/978-3-319-12206-9_4

Thériault, M., Claramunt, C., Séguin, A.-M., Villeneuve, P., 2002. Temporal GIS and Statistical Modelling of Personal Lifelines, in: Richardson, D.E., van Oosterom, P. (Eds.), *Advances in Spatial Data Handling*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 433–449. https://doi.org/10.1007/978-3-642-56094-1_32

Thériault, M., Séguin, A.-M., Aube, Y., Villeneuve, P.Y., 1999. A spatio-temporal data model for analysing personal biographies. *IEEE*, pp. 410–418. <https://doi.org/10.1109/DEXA.1999.795202>

Thomas, J.J., IEEE Computer Society (Eds.), 2005. *Illuminating the path: the research and development agenda for visual analytics*. IEEE Computer Soc, Los Alamitos, Calif.

Troncy, R., Shaw, R., Hardman, L., 2010. LODÉ: une ontologie pour représenter des événements dans le web de données, in: DESPRES, S. (Ed.), *21es Journées Francophones d'Ingénierie Des Connaissances (IC 2010)*, [Http://www.Ic2010.Mines-Ales.Fr/](http://www.Ic2010.Mines-Ales.Fr/). Ecole des Mines d'Alès, Nîmes, France, pp. 69–80.

Umiliacchi, P., Henning, U., Langer, G., Shingler, R., 2008. Standardized Data Interchange Between Railway Systems: an Integrated Railway Information System, in: *WCRR '08–World Conference on Railway Research–Seoul, Korea–18-22 May*.

Uschold, M., Gruninger, M., 1996. *Ontologies: Principles, methods and applications* M Uschold. *Knowl. Eng. Rev.* 11.

Vandersmissen, M.-H., Séguin, A.-M., Thériault, M., Claramunt, C., 2009. Modeling propensity to move after job change using event history analysis and temporal GIS. *J. Geogr. Syst.* 11, 37–65. <https://doi.org/10.1007/s10109-009-0076-x>

Villanova-Oliver, M., 2018. *Description du corpus de matériaux bruts (Livrable No. Livrable 1.1)*, ANR CHOUCAS. Laboratoire d'Informatique de Grenoble, Grenoble, France.

Villanova-Oliver, M., Boudis, M., Hombiat, A., Gensel, J., Le Rubrus, B., 2013. Analyse territoriale du bien-être avec HyperAtlas, in: INFORSID. Paris, France, pp. 149–164.

Villanova-Oliver, M., Noël, D., Gensel, J., Le Quéau, P., 2019. Un framework pour la description et l'analyse des trajectoires de vie : un pas vers l'étude des facteurs de migrations résidentielles, in: *Local Resources and Well-Being: A Multidisciplinary Perspective*.

Wang, H., Xinming, T., Bong, L., 2005. Modeling Spatial-temporal Data in Version-difference Model [WWW Document]. URL [/paper/Modeling-Spatial-temporal-Data-in-Model-Huibing-Xinming/284ca14fc835becdca827787d7e82bbb1b1e3dc8](http://paper/Modeling-Spatial-temporal-Data-in-Model-Huibing-Xinming/284ca14fc835becdca827787d7e82bbb1b1e3dc8) (accessed 7.18.18).

Wannous, R., Vincent, C., Malki, J., Bouju, A., 2015. An Ontology-Based Approach for Handling Explicit and Implicit Knowledge over Trajectories, in: Morzy, T., Valduriez, P., Bellatreche, L. (Eds.), *New Trends in Databases and Information Systems*. Springer International Publishing, Cham, pp. 403–413. https://doi.org/10.1007/978-3-319-23201-0_41

Watine, X., 2018. Optimisation de l'application Web ShinyDialect dédiée à la cartographie interactive de données géolinguistiques (Mémoire de Stage). Laboratoire d'Informatique de Grenoble, Grenoble, France.

Willekens, F.J., 1991. Understanding the Interdependence Between Parallel Careers, in: Siegers, J.J., de Jong-Gierveld, J., van Imhoff, E. (Eds.), *Female Labour Market Behaviour and Fertility*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11–31. https://doi.org/10.1007/978-3-642-76550-6_2

Ysebaert, R., Salmon, I., Le Rubrus, B., Bernard, C., 2014. Recueil, traçabilité et restitution des données territoriales du programme ESPON, in: *CIST2014 - Fronts et Frontières Des Sciences Du Territoire*. Collège international des sciences du territoire (CIST), Paris, France.

Zanin, C., Lambert, N., Grasland, C., Davoine, P.-A., Mathian, H., Saint-Marc, C., 2013. Recent Computer Technologies for an Innovative Cartographic Language (Technical Report No. 1). UMS Riate ; Laboratoire d'Informatique de Grenoble ; UMR Géographie-Cités.

Zheni, D., Frihida, A., Ghezala, H.B., Claramunt, C., 2009. A Semantic Approach for the Modeling of Trajectories in Space and Time, in: Heuser, C.A., Pernul, G. (Eds.), *Advances in Conceptual Modeling - Challenging Perspectives*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 347–356. https://doi.org/10.1007/978-3-642-04947-7_41

E ncadrement Doctoral

Thèses Soutenues (4)

- [TH-1] **1^{er} Oct. 2003 – 19 Déc. 2007** : **Ioan Marius Bilasco** *Une approche sémantique pour la réutilisation et l'adaptation de données 3D*.
Université Joseph Fourier, Grenoble
Financement Ministère de l'Enseignement Supérieur et de la Recherche.
Co-encadrement assuré à 50% et partagé avec H. Martin (Pr UPMF, 50%).
Durée de la thèse : 4 ans et 3 mois.
- [TH-2] **1^{er} Oct. 2006 – 8 Déc. 2009** : **Alina Dia Miron** *'Découverte d'associations sémantiques pour le Web sémantique géospatial : le framework ONTOAST'*.
Université Joseph Fourier, Grenoble.
Financement Ministère de l'Enseignement Supérieur et de la Recherche.
Co-encadrement assuré à 50% et partagé avec J. Gensel (Pr UPMF, 50%).
Durée de la thèse : 3 ans et 2 mois.
- [TH-3] **1^{er} Nov. 2012 – 24 Fév. 2017**: **Anthony Hombiat** *Un nouveau méta-modèle pour structurer la folksonomie d'OSM en une nouvelle ontologie*
Université Grenoble Alpes
Financement Université Pierre Mendès France et Minatec Idea's Laboratory
Co-encadrement assuré à 60% et partagé avec J. Gensel (Pr UGA, 40%).
Durée de la thèse : 4 ans et 3 mois
- [TH-4] **6 Jan. 2014 – 19 Mai 2017** : **Cécile Saint-Marc**, *Représentations cartographiques et géovisualisation des dynamiques spatio-temporelles des risques naturels : applications aux risques d'inondation impactant le réseau ferroviaire français*
Université Grenoble Alpes
Financement CIFRE-ANRT, en partenariat avec la SNCF
Co-encadrement assuré à 50% et partagé avec P-A. Davoine (Pr UGA, 50%)
Durée de la thèse : 3 ans et 4 mois

Thèses en cours (4)

- [TH-5] **1^{er} Oct. 2014 – Déc 2018** (prévision) : **David Noël**, *Trajectoires de vie multidimensionnelles pour l'analyse des migrations résidentielles au sein d'un espace métropolitain*.
Université Grenoble Alpes
Financement : Région Rhône-Alpes (ARC 7)
Co-encadrement assuré à 50% et partagé avec J. Gensel (Pr UGA, 40%) et Pierre Le Quéau (MCF-HDR, UGA/Pacte, 10%).
- [TH-6] **1^{er} Oct 2015 – Février 2019** (prévision) : **Camille Bernard**, *Gestion et visualisation spatio-temporelles de matrices d'interaction territoriale à l'heure du Web des Données*
Université Grenoble Alpes
Financement : Région Rhône-Alpes (ARC 7)
Co-encadrement assuré à 50% et partagé avec J. Gensel (Pr UGA, 40%) et Hy Dao (Pr. Université de Genève, 10%)
Remarque : interruption de 4 mois pour congés maternité en 2016.

- [TH-7] **1^{er} Mars 2017 – Février 2020** (prévision) : **Lina Toro**, *Visualisation et analyse spatiale des données de santé au travail à des fins de vigilance sanitaire*
 Université Grenoble Alpes
 Financement : École des Hautes Etudes en Santé Publique (contrat doctoral EHESP)
 Co-encadrement assuré à 50% et partagé avec V. Bonnetterre (Pr. UGA/TIMC, 50%)
- [TH-8] **1^{er} Oct. 2018 – 30 Sept. 2021** (prévision) : **Mathieu Viry**, *Géovisualisation pour l'aide à la décision dans un contexte d'incertitude : application à la localisation de victime en montagne*
 Université Grenoble Alpes
 Financement : ANR – Projet Choucas (ANR-16-CE23-0018)
 Co-encadrement assuré à 50% et partagé avec P-A. Davoine (Pr UGA, 50%)

Encadrement de Masters 2 Recherche

- 2018 : **Matthew Sreeves** (Master 2 Grenoble INP - MOSIG) sur le sujet '*Managing Uncertainty in Mountain Search and Rescue*'. Co-encadrement avec P-A. Davoine (Laboratoire Pacte)
- 2014 : **David Noël** (Master 2 UGA - ICA DCISS) sur le sujet '*Recitoire : Application client/serveur dédié à la collecte et l'analyse de données citoyennes*', réalisé dans le cadre du projet de recherche inter SFR FabTer.
- 2012 : **Soreangsey Kiv** (Master 2R SIGAL UJF) sur le sujet '*Approche méthodologique pour l'aide à la conception d'environnements de géovisualisation*'. Co-encadrement avec P-A. Davoine (équipe Steamer/LIG)
- 2010 : **Qiang Fu** (Master 2R MoSIG UJF) sur le sujet '*Une approche sémantique pour la détection d'incohérences entre informations provenant du web*'. Co-encadrement avec M-C Rousset (équipe HADAS/LIG)
- 2010 : **Aastha Sharma** (Master 2R MoSIG UJF) sur le sujet '*Semantic Web Services Description and Classification*' Co-encadrement (avec D. Ziébelin, équipe Steamer/LIG)
- 2007 : **Ramon Lamana** (Master 2R Intelligence, Interaction et Information UJF) sur le sujet '*Approche déclarative pour la réutilisation de contenus 3D*'. Encadrement assisté de I.M. Bilasco (équipe Steamer/LIG)
- 2004 : **Ba-Quan Nguyen** (Master 2R Intelligence, Interaction et Information UJF) sur le sujet '*Adaptativité dans les Systèmes d'Information sur le Web*'. Co-encadrement avec J. Gensel (équipe Steamer/LIG)
- 2003 : **Ioan Marius Bilasco** (DEA Informatique, Système et Communication de l'Université Joseph Fourier) sur le sujet '*Modèles de Présentations Multimédias Adaptables*'. Co-encadrement avec J. Gensel (équipe Steamer/LIG).

Représentations de connaissances spatiales évolutives : des ontologies aux géovisualisations

Habilitation à Diriger des Recherches

Marlène Villanova-Oliver

Résumé :

Au cours de la dernière décennie, le contexte de nos recherches s'est vu modelé par des avancées scientifiques et techniques autour du Web Sémantique, par des changements sociétaux importants dont l'essor des sciences participatives et par des initiatives émanant de différentes instances de gouvernance, promouvant des directives telles qu'Inspire ou encore la démarche Open Data. Ces éléments, entre autres, contribuent à une mise à disposition sans précédent de données, d'informations et de connaissances qui peuvent être mobilisées dans la compréhension de phénomènes (de nature sociodémographique, environnementale, sociologique, etc.) et dans la prise de décision. Les recherches décrites dans ce document abordent des questions relatives *aux* représentations, conceptuelle d'une part et géovisuelle d'autre part, des connaissances immergées dans le Web des Données Ouvertes et Liées.

Le cadrage scientifique des travaux présentés est ainsi donné par le champ de la représentation des connaissances, ici à base d'ontologies versées dans le Web Sémantique et, par le champ de la géovisualisation, appliqués à des connaissances spatiales évolutives, potentiellement porteuses d'imperfection et exploitées à des fins de compréhension de phénomènes et/ou de prise de décision. Trois approches sont présentées, issues de thèses qui ont débuté à partir de 2014. Elles s'inscrivent chacune dans un contexte pluridisciplinaire particulier, et s'organisent autour d'objets d'étude pour lesquels la notion même d'évolution prend des formes différentes. Les réalisations proposées répondent à trois objectifs : *i)* garantir une certaine genericité des approches bien que les ontologies développées soient des solutions à des problèmes concrets, *ii)* intégrer dimension descriptive et dimension explicative qui favorisent selon nous la compréhension des phénomènes et la prise de décision, *iii)* supporter l'imperfection potentiellement inhérente aux informations manipulées. Nous abordons ainsi successivement : la gestion des différentes versions de découpages territoriaux et la documentation des changements survenus, l'outillage méthodologique et logiciel pour l'étude des trajectoires de vies, l'enrichissement des approches de type Retour d'Expérience par le récit géovisuel de situations de crise.

Des perspectives de recherche, nourries par ces travaux, mettent l'accent sur le lien entre représentation de connaissances, géovisualisation et raisonnement humain que nous proposons d'explorer plus finement pour mieux répondre aux besoins de l'utilisateur en situation de prise de décision.
